Quantitative amyloid PET imaging in Alzheimer's disease

Heeman, Fiona

2021

document version
Publisher's PDF, also known as Version of record

Quantitative amyloid PET imaging in Alzheimer's disease

Walking the line between accuracy and simplicity

Fiona Heeman

2021

The research described in this thesis was carried out at the department of Radiology & Nuclear Medicine from the Amsterdam UMC, VUmc, Amsterdam, the Netherlands. The project leading to this thesis has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115952. This Joint Undertaking receives the support from the European Union's Horizon 2020 research and innovation programme and EFPIA http://www.imi.europa.eu. This communication reflects the views of the authors and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. Printing of this thesis was financially supported by GE Healthcare, Life Molecular Imaging GmbH, Alzheimer Nederland and the Graduate School Neurosciences Amsterdam Rotterdam (ONWAR).

ISBN: 978-94-6421-482-6
Cover: Wendy Bour, Ipskamp Printing & Fiona Heeman
Lay-out & Design: Yasmin Katlich, Persoonlijkproefschriften
Printing: Ipskamp Printing, Enschede

VRIJE UNIVERSITEIT

Quantitative amyloid PET imaging in Alzheimer's disease

Walking the line between accuracy and simplicity

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. C.M. van Praag,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op woensdag 3 november 2021 om 13.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Fiona Heeman

geboren te Dronten

promotor:	prof.dr. A. A. Lammertsma
copromotoren:	dr.ir. M.M. Yaqub
	dr. I. Lopes Alves

Leescommissie:	prof.dr. F. Barkhof
	prof.dr. M. Koole
	prof.dr. G. Moos Knudsen
	dr. R. Ossenkoppele
	dr. V. Garribotto
	dr. D. van Assema
Paranimfen:	J.J. Eertink
	M.K.A.C. van der Werf

The important thing is to not stop questioning. Curiosity has its own reason for existing

– Albert Einstein

Content

1.

General introduction

Alzheimer's disease dementia

Globally, the ever increasing age of the population has led to an increased incidence of dementia (*1*). In the Netherlands alone, there are currently over 280.000 people suffering from dementia and the risk of developing dementia during life has reached 20% (*2*). Its impact is two-fold: on one hand, the disease places an enormous burden on the patient, family and caregivers, and on the other, it constitutes a large financial burden for society as a whole (*3*,*4*). More specifically, it has been estimated that in 2017, the costs of dementia care for the Dutch healthcare system were approximately 9.1 billion Euros, i.e. 10.3% of the total Dutch healthcare costs (*5*).

At present, the most common cause of dementia is Alzheimer's disease (AD), which comprises approximately 70% of all cases (*2*,*4*,*6*). AD dementia is characterised by severe cognitive deterioration, which interferes with patients' daily life functioning (*4*,*7*). Clinical symptoms typically include memory impairment and problems with other cognitive abilities such as executive and visual-spatial functions, or language (*7*). In addition, possible neuropsychiatric and behavioural changes may co-occur (*7*–*9*).

Pathophysiology of Alzheimer's disease

The first pathological changes of AD occur decades before clinical onset of dementia (*10*). These changes typically include accumulation of extracellular amyloid-beta (Aβ) plaques (also known as neuritic plaques) and intracellular hyperphosphorylated neurofibrillary tau tangles, resulting in atrophy (i.e. loss of brain tissue) (*10*). The accumulation of Aβ plaques is hypothesised to be one of the earliest changes in the brain, initiating downstream biological processes such as tau aggregation, neuro-inflammation and synaptic failure, eventually leading to neuronal loss and cognitive impairment. This widely accepted hypothesis is known as the amyloid cascade hypothesis (*11*). By placing accumulation of Aβ plaques at the beginning of the disease process, this hypothesis suggests that these plaques should be an ideal biomarker (i.e. indicator of a biological process) for AD. This is illustrated by the recent introduction of positron emission tomography (PET) tracers for measuring Aβ plaques, which has led to an increased confidence in diagnosing AD dementia during life (*12*).

Positron Emission Tomography

PET is a nuclear medicine imaging technique that allows for *in vivo* visualisation and quantification of functional and molecular processes, such as the accumulation of Aβ plaques, using an appropriate radiotracer (*13*,*14*). Such a radiotracer is a positron

emitting radionuclide (such as fluorine-18 or carbon-11) bound to a ligand that interacts with the target of interest, in this case the Aβ plaques (*14*). Due to its unstable nature, the radionuclide emits positrons, which travel a short distance (at most 1-2 mm) before colliding with an electron. This collision results in the annihilation of both particles and the emission of two gamma rays (511 *ke*V each) in opposite directions (Figure 1) (*15*). The simultaneous detection of these gamma rays by opposite PET detectors (coincidence detection) allows for localization of the line of response, which connects the two detectors. Measurement of sufficient lines of response will enable the reconstruction of a 3D image of the distribution of tracer radioactivity within the field of view (FOV). This distribution will be focussed on the target of interest, where the vast majority of tracer is bound. Furthermore, combining multiple (consecutive) reconstructions over time allows for capturing the process of tracer distribution in a 4D image. Finally, provided that an appropriate acquisition protocol is used, quantitative estimates of this distribution can be derived through application of mathematical models, which will be discussed in more detail later.

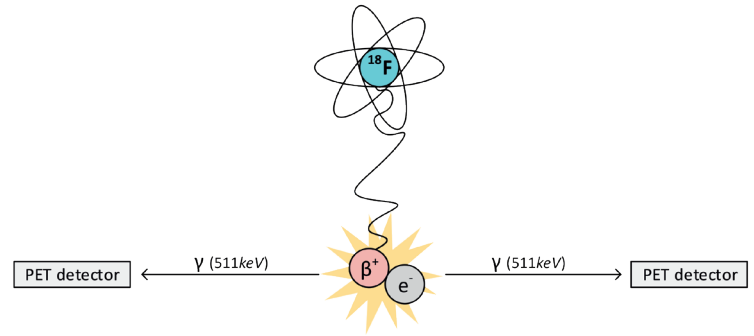


Figure 1. Schematic illustration of the principle of PET detection
Annihilation of a positron (β^+) and an electron (e^-) resulting in the simultaneous emission of two gamma (γ) rays, which can be detected by in coincidence detectors. ^{18}F: fluorine-18.

Amyloid PET tracers

In 2004, one of the first successful PET tracers for measuring Aβ plaques, carbon-11 labelled Pittsburgh compound B ([^{11}C]PiB), was developed. After successfully validating results obtained with [^{11}C]PiB PET against neuropathological findings, this tracer could be introduced in clinical routine (*16*). However, as carbon-11 has a half-life (i.e. the time it takes for the total amount of radioactivity to reduce by a factor of two) of only 20

minutes, this tracer can only be used by centres with an on-site cyclotron and associated radiochemistry facilities. To overcome this limitation, novel PET tracers targeting Aβ plaques were developed and labelled with fluorine-18 (F-18), which has a half-life of 110 minutes. The longer half-life of fluorine-18 facilitates off-site usage, enabling transport of these tracers to centres without a cyclotron facility, making them commercially viable. These commercially available tracers, such as [^{18}F]florbetapir, [^{18}F]florbetaben and [^{18}F]flutemetamol (*17–21*), are able to identify Aβ plaques with high sensitivity and specificity, as shown by post-mortem studies (*16*,*22–25*), hence they were officially approved for clinical use shortly after their introduction.

PET imaging of Aβ plaques

With the introduction of radiotracers targeting Aβ plaques, amyloid PET imaging has recently been implemented in the diagnostic work-up of dementia patients. For this (diagnostic) application, PET scans are used to visually assess whether Aβ plaques are present or not (i.e. a binary classification of the scan). Here, the lack of Aβ plaques strongly increases a physician's confidence in ruling out AD as cause of cognitive symptoms (*26*). On the other hand, the presence of these Aβ plaques does not necessarily result in a conclusive diagnosis, as they are neither specific for, nor sufficient to characterise AD (*26*), as presence of Aβ plaques has also been reported in asymptomatic individuals. To date, it remains unclear whether these plaques have a causal relationship with AD dementia. In line with pathological studies (*27*), recent work has shown that various stages of regional Aβ plaque pathology can also be detected with PET. These stages precede global positivity and appear to be specific for the early disease phases (*28*). Together, this emphasizes the need for better understanding of the role of the Aβ plaques in the development of the disease, which can be achieved by accurately monitoring its involvement throughout the course of AD. By quantitatively tracking Aβ plaques using PET, it is possible to obtain detailed information on amount, location and accumulation rates. Having this information may 1) enhance our understanding of AD, 2) improve our ability to predict who will develop AD dementia, and 3) allow for a better selection of participants for secondary prevention trials. Eventually, once a successful intervention has been developed, PET could contribute to patient selection and response monitoring in a clinical setting.

PET acquisition

To obtain highly accurate estimate of Aβ pathology, a (long) dynamic (time-sequence) scan is required (*14*). A dynamic scan takes into account physiological processes

involved in the uptake of a tracer, such as cerebral blood flow (CBF), tracer extraction and clearance. In combination with the measurement of available tracer in plasma, this allows for estimating the amount of tracer in tissue that is specifically bound to the target, in this case the Aβ plaques. As an alternative to a dynamic scan, a short static scan can be acquired at tracer equilibrium conditions. This scan provides only a "snapshot" of the tracers' target engagement, as it cannot account for other physiological processes (*29*). Clearly, the first type of scan provides more accurate measurements, but it is also more complex in terms of acquisition and analysis protocols. On the other hand, a static scan offers a much simpler procedure and requires only simple calculations, but it may be biased due to multiple other physiological processes (*30*,*31*).

Pharmacokinetic modelling

To understand how measured radioactivity in plasma together with a dynamic PET scan can provide quantitative estimates of Aβ pathology requires a basic understanding of the concept of pharmacokinetic modelling. The tracer distribution over time that is measured by the PET scanner consists of a combination of signals stemming from distinct physiological processes and only one of them needs to be extracted (*14*), i.e. tracer specifically bound to the Aβ plaques. In order to do so, a simplification of the biological complexity is required by describing kinetics of the tracer and its various states using a mathematical model (*14*). To this end, a compartmental approach is adopted, which describes uptake of the tracer in different compartments (states) as a function of time (*14*). These compartments are not necessarily physical entities, but rather states in time that assume homogenous tracer concentrations. Examples of such compartments are freely diffusible tracer, either in plasma or in tissue and specifically or non-specifically bound tracer in tissue (Figure 2). Initial analyses of a novel tracer aim to determine which compartmental model best describes its kinetics. Next, this model is used to extract the signal of interest from the PET image. Based on the input (i.e. measured tracer concentrations in plasma) and output (i.e. PET image) the optimal model fit can be determined, i.e. the fit that results in the smallest differences between estimated (model) and measured PET data. The resulting fit parameters describe the transfer rates between compartments and are referred to as kinetic parameters or rate constants (Figure 2). Ratios of these rate constants can be used to describe tracer binding in the tissue compartment (i.e. $K_1/k_2 \cdot (1+ k_3/k_4)$ = volume of distribution, V_T and k_3/k_4 = non-displaceable binding potential, BP_{ND}) (*32*). It should be noted that depending on the tracer and under specific conditions (i.e. negligible variation in the extraction fraction), the delivery rate constant K_1 can be used as proxy for CBF ($K_1 = Extraction \cdot CBF$) (*14*).

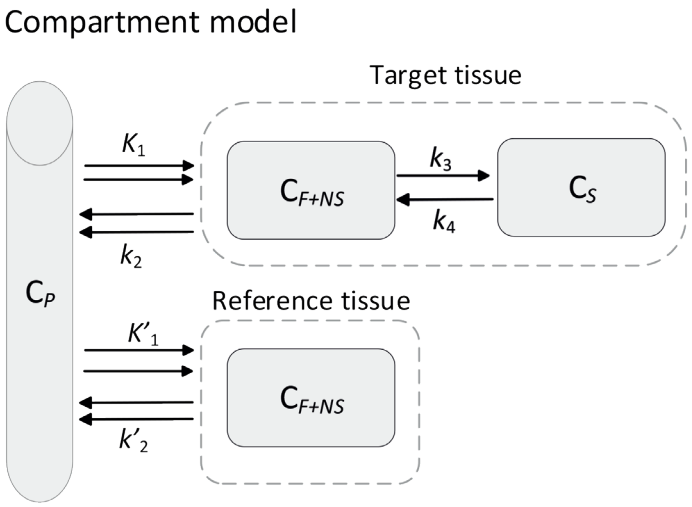


Figure 2. Schematic overview of a compartment model C_p: tracer radioactivity concentration in plasma, C_{F+NS}: free and non-specifically bound concentration in tissue, C_S: specifically bound concentration in tissue, all units Bq·mL^{-1}. Rate constants (i.e. *K*'s) describe the transfer from one compartment to the other, units of the rate constants are: K_1 in mL·g^{-1}·min^{-1} and k_2, k_3, k_4 in min^{-1}.

Simplifications and assumptions

Full dynamic quantification and pharmacokinetic modelling are time consuming and complex procedures; hence, many efforts have been undertaken to simplify them. Simplified methods, however, rely on certain assumptions that may not be true under all conditions. Considering that each tracer has a distinct kinetic behaviour, these simplifications need to be validated for each tracer separately.

A routinely used simplification is the reference tissue approach, which replaces the need for arterial blood sampling by providing an alternative input function (i.e. the tracers' activity over time in a reference tissue). A reference tissue is a region within the brain that is devoid of specific binding sites, but otherwise has similar tissue characteristics as the target tissue (i.e. similar non-specific binding) and tracer uptake that is unaffected by the disease under investigation (*33*). This approach allows for kinetic modelling to derive estimates of the specifically bound tracer (BP_{ND}) and relative tracer delivery rate (R_1) which is the ratio of target (K_1) to reference tissue delivery rate (K_1'), Figure 2. Various pharmacokinetic models that use a reference tissue approach have been developed, such as the simplified reference tissue model (SRTM), the multilinear reference tissue model (MRTM) and reference Logan (RLogan) (*33*–*35*), and each of

these models has a unique set of assumptions. For example, SRTM assumes that both target and reference tissue kinetics are best described by a single compartment, that there is negligible contribution of intravascular activity to both target and reference tissues and no displaceable component in the reference tissue (*33*). On the other hand, linear models such as MRTM and RLogan assume there is a certain point in time where the target to reference tissue concentration is and remains constant (*34*,*36*). For analysis of very noisy data, corresponding to single PET voxels (i.e. three-dimensional "pixels" of a PET image), specific implementations of these models have been developed. These are called parametric models and include the basis function approach of SRTM, receptor parametric mapping (RPM), and SRTM and MRTM with fixed k_2' parameter, known as SRTM2 and MRTM2, respectively.

Another routinely used simplification is the adoption of static acquisition protocols, which assume an equilibrium between target and reference tissue tracer concentrations, and the absence of changes in physiological processes such as cerebral blood flow, tracer extraction or clearance in the case of consecutive scans (*37*). From this scan, the semi-quantitative standardized uptake value ratio (SUVR) can be calculated, which is the tracer concentration in the target tissue relative to that in a reference tissue. In this approach, it is assumed that changes in SUVR signify changes in specific binding, as opposed to changes in other physiological processes. Moreover, it implies that SUVR-1 would be comparable with BP_{ND} derived from a dynamic scan. Nonetheless, from a static scan, no distinction can be made between freely available tracer in tissue, and specifically and non-specifically bound tracer. Thus, if changes in underlying physiological processes occur, SUVR will be a biased proxy of specific binding (*29*,*30*,*38*). Furthermore, it has been demonstrated that SUVR may be more variable than reference tissue model derived BP_{ND}, which suggests that BP_{ND} is more suitable for measuring small changes (*39*). Recently, another alternative to the full dynamic scanning protocol has been introduced, namely a dual-phase or dual-time window protocol (*40*,*41*). This protocol consists of two shortened scanning phases, which significantly reduces total scan time compared with a standard dynamic scan, but which still allows for kinetic modelling to derive estimates of the specifically bound tracer (BP_{ND}) and relative tracer delivery rate (R_1).

Challenges of simplified methods

In nuclear medicine, simplified procedures often are favoured over complex, full quantitative procedures as they are more easily implemented, especially in clinical routine. Unfortunately, this also means that these simplifications are sometimes

implemented without proper validation against the gold standard, i.e. full quantification using a dynamic scan and plasma input function. Nonetheless, especially for novel tracers, it remains crucial to assess whether the assumptions of simplified procedures have been met and to understand the consequences of full or partial violation (*42*). This is particularly important because improper use of simplifications may result in poor quantitative accuracy, which could lead to erroneous conclusions. Thus, thorough validation is required before choosing simplified approaches.

Factors that may compromise the use of simplified approaches can be either assumption violations or implementation errors. In practise, the reference tissue model assumption of similar reference and target tissue characteristics is frequently violated due to the use of regions that partially or fully consist of a different tissue (e.g. white matter rather than grey matter). This violation may compromise measurement accuracy for both quantitative and semi-quantitative approaches. Furthermore, there may be changes in physiological processes such as tracer extraction, clearance or CBF, as a result of disease-progression or as side effects of a specific intervention (*43*). If these changes occur between consecutive scanning sessions, they may affect the semi-quantitative metric SUVR, regardless of changes in specific binding. In addition, semi-quantitative approaches may also suffer from implementation errors such as delayed or expedited scan start times, which could interfere with the equilibrium assumption (*31*).

Finally, it should be noted that, as for full quantitative procedures, simplified procedures also require a thorough understanding of the precision of the measurement, to allow for making reliable statements regarding the process of interest, e.g. the observed change in Aβ pathology.

AMYPAD- Amyloid imaging to prevent Alzheimer's disease

With the widespread application of amyloid PET imaging, several questions regarding its clinical and predictive value remain. Aiming to address some of these issues, the international, collaborative research initiative AMYPAD aims to improve understanding, diagnosis and management of AD by collecting large quantities of clinical data and acquiring amyloid PET scans using two of the recently introduced tracers: [^{18}F]florbetaben and [^{18}F]flutemetamol (*44*). The project consists of two studies: the Diagnostic and Patient Management Study (DPMS) and the Prognostic and Natural History Study (PNHS) (*45*,*46*). The DPMS focusses on the use of amyloid PET scans in clinical routine, where the presence or absence of Aβ plaques is assessed visually by a

nuclear medicine physician. In particular, this study aims to determine the value of this assessment for both diagnosis and patient management. On the other hand, the PNHS aims to track and predict disease progression with amyloid PET in early stages of AD. Therefore, more fine-grained estimates of Aβ pathology are required, which can be obtained via quantification. Hence, most of the work presented in this thesis relates to the AMYPAD PNHS.

To date, for each of the PET tracers measuring Aβ plaques, implementations of simplified procedures have been validated to a limited extent. Consequently, knowledge gaps remain, which need to be investigated for a full understanding of the applicability of simplified approaches. These gaps are related to technical advances, such as the implementation of novel, dual-time window acquisition protocols, as well as to the shifted focus in the field of amyloid imaging in AD from clinical to early disease stages, (i.e. before onset of clinical symptoms) and from global to regional measures of Aβ pathology.

Thesis aims

This thesis aims to evaluate to what extent approaches for quantification of Aβ pathology can be simplified, within the context of different applications. Of course, a perfect balance between accuracy and simplicity cannot be determined in isolation, but needs to be considered in the light of a specific clinical or research question. Ultimately, this knowledge will enable informed decision making on whether simplifications are appropriate. Hence, the goals of this thesis were the following:

Quantification of amyloid PET scans

First, **Chapter 2** describes various strategies for Aβ quantification implemented in AMYPAD and the efforts made to improve comparability between results from [^{18}F]flutemetamol and [^{18}F]florbetaben tracers. **Chapter 3** aims to determine the optimal dual-time window scanning protocol by means of a simulation study. **Chapter 4** extends this work by validating the optimal parametric methods for analyses of dual-time window scans. Next, **Chapter 5** evaluates variability of [^{11}C]PiB-derived R_1 using test-retest data.

Impact of confounding factors

Chapter 6 assesses the impact of simulated changes in CBF on SUVR and DVR, for both [^{18}F]flutemetamol and [^{18}F]florbetaben and evaluates the effect of varying acquisition start times on SUVR. Next, **Chapter 7** builds onto this simulation work, by comparing SUVR

with DVR to understand the impact of CBF and extent of Aβ burden on bias observed in SUVR relative to DVR. Last, **Chapter 8** examines the effect of using alternative reference tissues for normalization purposes on full and semi-quantitative approaches for [^{11}C]PiB.

Impact on clinical practise and trials

Chapter 9 introduces a regional multi-tracer model for staging cortical Aβ and evaluates its potential for predicting cognitive decline. Finally, **Chapter 10** investigates strategies for reducing patient sample sizes in clinical trials, including choice of acquisition protocol and use of regional or global measures of Aβ burden.

References

1. Winblad B, Amouyel P, Andrieu S, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. The Lancet Neurology. 2016;15.
2. Ott A, Breteler MMB, Birkenhäger-Gillesse EB, Harskamp, van F, Koning, de I, Hofman A. De prevalentie bij ouderen van de ziekte van Alzheimer, vasculaire dementie en dementie bij de ziekte van Parkinson; het ERGO-onderzoek. Nederlands Tijdschrift voor Geneeskunde. January 1996:200-205.
3. Prince M, Wimo A, Guerchet M, Ali G-C, Wu Y-T, Prina M. World Alzheimer Report 2015. The Global Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends.; 2015.
4. 2016 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2016;12:459-509.
5. RIVM. Zorguitgaven dementie 2017. 2017.
6. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. Lancet. 2016;388:505-517.
7. Association AP. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013.
8. Mega MS, Cummings JL, Fiorello T, Gornbein J. The spectrum of behavioral changes in Alzheimer's disease. Neurology. 1996;46:130-135.
9. Zhao Q-F, Tan L, Wang H-F, et al. The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis. J Affect Disord. 2016;190:264-271.
10. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010;9:119-128.
11. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. Science. 1992;256:184-186.
12. de Wilde A, van der Flier WM, Pelkmans W, et al. Association of Amyloid Positron Emission Tomography With Changes in Diagnosis and Patient Treatment in an Unselected Memory Clinic Cohort: The ABIDE Project. JAMA Neurol. 2018;75:1062-1070.
13. Ter-Pogossian MM, Phelps ME, Hoffman EJ, Mullani NA. A Positron-Emission Transaxial Tomograph for Nuclear Imaging (PETT). Radiology. 1975;114:89-98.
14. Morris ED, Endres CJ, Schmidt KC, Christian BT, Muzic RF, Fisher RE. Kinetic Modeling in Positron Emission Tomography. Emission Tomography: The Fundamentals of PET and SPECT. November 2004:499-540.
15. Saha GB. Basics of PET imaging: physics, chemistry, and regulations. New York, NY: Springer; 2005.
16. Ikonomovic MD, Klunk WE, Abrahamson EE, et al. Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain. 2008;131:1630-1645.

17. Koole M, Lewis DM, Buckley C, et al. Whole-Body Biodistribution and Radiation Dosimetry of 18F-GE067: A Radioligand for In Vivo Brain Amyloid Imaging. Journal of Nuclear Medicine. 2009;50:818-822.
18. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 Study of the Pittsburgh Compound B Derivative 18F-Flutemetamol in Healthy Volunteers and Patients with Probable Alzheimer Disease. Journal of Nuclear Medicine. 2009;50:1251-1259.
19. Lin K-J, Hsu W-C, Hsiao I-T, et al. Whole-body biodistribution and brain PET imaging with [18F]AV-45, a novel amyloid imaging agent--a pilot study. Nucl Med Biol. 2010;37:497-508.
20. Wong DF, Rosenberg PB, Zhou Y, et al. In Vivo Imaging of Amyloid Deposition in Alzheimer's Disease using the Novel Radioligand [18F]AV-45 (Florbetapir F 18). J Nucl Med. 2010;51:913-920.
21. Rowe CC, Ackerman U, Browne W, et al. Imaging of amyloid β in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. The Lancet Neurology. 2008;7:129-135.
22. Morris E, Chalkidou A, Hammers A, Peacock J, Summers J, Keevil S. Diagnostic accuracy of (18)F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2016;43:374-385.
23. Ikonomovic MD, Buckley CJ, Heurling K, et al. Post-mortem histopathology underlying β-amyloid PET imaging following flutemetamol F 18 injection. acta neuropathol commun. 2016;4:130.
24. Clark CM, Pontecorvo MJ, Beach TG, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study. Lancet Neurol. 2012;11:669-678.
25. Sabri O, Sabbagh MN, Seibyl J, et al. Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: Phase 3 study. Alzheimer's & Dementia. 2015;11:964-974.
26. Villemagne VL, Barkhof F, Garibotto V, Landau SM, Nordberg A, van Berckel BNM. Molecular Imaging Approaches in Dementia. Radiology. January 2021:200028.
27. Thal DR, Rüb U, Orantes M, Braak H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. Neurology. 2002;58:1791-1800.
28. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. Neurology. 2017;89:2031-2038.
29. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. J Nucl Med. 2017;58:1019-1024.
30. Berckel van BNM, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. J Nucl Med. 2013;54:1570-1576.

31. Schmidt ME, Chiao P, Klein G, et al. The influence of biological and technical factors on quantitative analysis of amyloid PET: Points to consider and recommendations for controlling variability in longitudinal data. Alzheimers Dement. 2015;11:1050-1068.
32. Lammertsma AA. Compartmental Modeling in Emission Tomography. In: Fleck I, Titov M, Grupen C, Buvat I, eds. Handbook of Particle Detection and Imaging. Cham: Springer International Publishing; 2020:1-20.
33. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996;4:153-158.
34. Ichise M, Liow J-S, Lu J-Q, et al. Linearized Reference Tissue Parametric Imaging Methods: Application to [11 C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human Brain. Journal of Cerebral Blood Flow & Metabolism. 2003;23:1096-1112.
35. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. J Cereb Blood Flow Metab. 1996;16:834-840.
36. Logan J, Fowler JS, Volkow ND, et al. Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects. J Cereb Blood Flow Metab. 1990;10:740-747.
37. Carson RE, Channing MA, Blasberg RG, et al. Comparison of Bolus and Infusion Methods for Receptor Quantitation: Application to [18 F]Cyclofoxy and Positron Emission Tomography. Journal of Cerebral Blood Flow & Metabolism. 1993;13:24-42.
38. Wolfensberger SP, Maruyama K, van Berckel BN, et al. Quantification of the neurokinin 1 receptor ligand [11C]R116301. Nucl Med Commun. 2011;32:896-902.
39. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008;42:76-86.
40. Bullich S, Barthel H, Koglin N, et al. Validation of Noninvasive Tracer Kinetic Analysis of 18F-Florbetaben PET Using a Dual-Time-Window Acquisition Protocol. J Nucl Med. 2018;59:1104-1110.
41. Cecchin D, Barthel H, Poggiali D, et al. A new integrated dual time-point amyloid PET/MRI data analysis method. European Journal of Nuclear Medicine and Molecular Imaging. 2017;44:2060-2072.
42. Salinas CA, Searle GE, Gunn RN. The simplified reference tissue model: model assumption violations and their impact on binding potential. J Cereb Blood Flow Metab. 2015;35:304-311.
43. Postiglione A, Lassen NA, Holman BL. Cerebral blood flow in patients with dementia of Alzheimer's type. Aging Clin Exp Res. 1993;5:19-26.

44. Farrar G, Barkhof F, Pla S, et al. AMYPAD: A EUROPEAN PUBLIC-PRIVATE PARTNERSHIP TO INVESTIGATE THE VALUE OF β-AMYLOID BRAIN SCANS AS A DIAGNOSTIC AND THERAPEUTIC MARKER FOR ALZHEIMER'S DISEASE. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2017;13:P420.
45. Lopes Alves I, Collij LE, Altomare D, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. Alzheimer's Dement. April 2020.
46. Frisoni GB, Barkhof F, Altomare D, et al. AMYPAD Diagnostic and Patient Management Study: Rationale and design. Alzheimers Dement. 2019;15:388-399.

PART I

Quantification of amyloid PET scans

2.

β-amyloid PET Imaging in the AMYPAD Project: Implementing Quantitative Imaging in Clinical Trials and Clinical Practice

Juan Domingo Gispert,* Isadora Lopes Alves,* Christopher Buckley, **Fiona Heeman**, Santiago Bullich, Gemma Salvadó, Maqsood Yaqub, Mahnaz Shekari, Daniele Altomare, Christopher Foley, Katherine R. Gray, Alessandro Palombit, Richard Manber, Robin Wolz, Derek L. Hill, Anja Mett, Lyduine E. Collij, Pawel Markiewicz, Jieqing Jiao, Mark Battle, Nigel Banton, M. Jorge Cardoso, Laure Saint-Aubert, Ronald Boellaard, Valentina Garibotto, Bart N.M. van Berckel, José Luis Molinuevo, Giovanni B. Frisoni, Andrew Stephens, Gill Farrar, Mark Schmidt, Adriaan A. Lammertsma, Frederik Barkhof
On behalf of the AMYPAD Consortium

Submitted

Abstract

Beta-amyloid (Aβ) deposition is an early pathogenic step on the path towards the development of Alzheimer's disease (AD). Understanding the value of imaging of Aβ using positron emission tomography (PET) provides a unique opportunity to achieve three major goals: 1) to improve the diagnostic workup of patients suspected to have AD and their management; 2) to understand the natural history of AD at pre-symptomatic and early symptomatic stages to better select patients for trials; and 3) to monitor changes in Aβ deposition to quantify the impact of novel therapies. The AMYloid imaging to Prevent Alzheimer's Disease (AMYPAD) project addresses the above goals. To this end, two multicentre clinical trials been initiated: the Diagnostic and Patient Management Study (DPMS) and the Prognostic and Natural History Study (PNHS).

This paper presents the approaches used to provide quantitative estimates of Aβ load in these two trials. Quantification pipelines based on the Centiloid principle have been set up to render comparable outcome measures, irrespective of whether scans were acquired using [^{18}F]florbetaben and/or [^{18}F]flutemetamol. For dynamic acquisitions, a dual-phase scanning protocol has been developed to balance the benefits and trade-offs for repeated dynamic scanning. Along with answers to relevant clinical and research questions, AMYPAD is expected to provide technical contributions to improve the quantitative accuracy, sensitivity and robustness of multi-centre, multi-tracer Aβ PET imaging for research purposes. For clinical use, it will expand the use of quantification as an aid to visual reading in diagnostic settings. Finally, AMYPAD PNHS will constitute a unique resource of dynamic Aβ PET scans to identify use cases where dynamic imaging and kinetic modelling might be superior to ratio metrics calculated on static acquisitions.

Keywords: Alzheimer's, clinical trials, multicentre, prevention, dynamic PET.

Introduction

Amyloid imaging to prevent Alzheimer's disease (AMYPAD; https://amypad.eu) is a collaborative research program to improve understanding, diagnosis and management of Alzheimer's disease (AD) through the utilisation of amyloid-β (Aβ) positron emission tomography (PET) imaging. AMYPAD is part of the Innovative Medicines Initiative, a joint undertaking between the European Commission and the European Federation of Pharmaceutical Industries and Associations. AMYPAD has three main goals: 1) to improve the diagnostic work-up and management of people suspected to have Alzheimer's disease, 2) to better understand the natural history of the disease in a pre-symptomatic stage and 3) to model disease progression to better inform treatment trials aiming at preventing AD. We expect a total of 4000-5000 scans to be performed and analysed in the AMYPAD project, equally split between the PET imaging agents [^{18}F]florbetaben and [^{18}F]flutemetamol. To achieve its goals, two multicentre clinical trials have initiated, i.e. the Diagnostic and Patient Management Study (DPMS) (*1*), and the Prognostic and Natural History Study (PNHS) (*2*).

The DPMS involves 8 centres and recently completed its enrolment. It selected and is following up a memory clinic population suspected of AD to determine the utility of Aβ PET imaging with regards to change in diagnosis, diagnostic confidence, and pharmacological and non-pharmacological management (*1*). Study populations are patients with subjective cognitive decline associated with features that increase the likelihood of preclinical AD (SCD Plus; *N*=245), mild cognitive impairment (MCI; *N*=341), and dementia (*N*=258). Participants are recruited at 8 European sites: Amsterdam University Medical Center (location VUmc) – Amsterdam (The Netherlands), Barcelonaβeta Brain Research Center – Barcelona (Spain), University of Cologne – Cologne (Germany), Cantonal Hospitals of Geneva and Lausanne (Switzerland), University College London – London (United Kingdom), Hôpitaux de Toulouse (France) and Karolinska Institutet – Stockholm (Sweden). In the DPMS, Aβ burden will be primarily defined based on visual reading, as per standard clinical routine. Quantification of Aβ burden will be performed for secondary analyses.

The PNHS involves 17 sites, which together aim to enrol 2000 non-demented participants (*2*). The AMYPAD PNHS is a natural history study that will evaluate how Aβ imaging might help to improve both understandings of the natural course of AD and modelling of biomarker trajectories along preclinical stages, with a special focus on enrolling subjects

with amyloid levels near traditional thresholds of biomarker abnormality (so-called grey-zone of amyloidosis). The study will use Aβ PET imaging as an additional and relevant AD biomarker to complement the phenotypical characterization of participants in several European parent study cohorts, including the EPAD LCS (*3*), EMIF-AD Twin 60++ (*4*), FACEHBI (*5*), ALFA+ (*6*), FPACK (*7*), and UCL-2010-412 cohort (*8*), with several others expected to join. In addition to prospective Aβ PET collection, a unique characteristic of the PNHS is the collection of historical scans from several longitudinal parent studies across Europe. Due to its research and early population focus, the primary measure of brain Aβ load in the PNHS will be quantitative both at baseline and all subsequent visits.

Considering the above objectives, this paper reports the challenges and adopted solutions for acquisition, analysis, and management of Aβ PET in both AMYPAD studies and presents early results. Also, expected contributions of the AMYPAD trials to the development of quantitative approaches for Aβ PET are discussed.

Methods (Analysis of Requirements)

Outcome measurements of Aβ burden

Although both AMYPAD studies focus on the use of Aβ PET data, their goals and therefore respective image assessment methods are inherently distinct. Table 1 shows a summary of the main characteristics of each of the trials.

As the DPMS is designed from a clinical-routine perspective, PET images in that study are visually read by local nuclear medicine physicians, whereas semi-quantitative metrics are only used as secondary outcome measures. Indeed, the study was designed to interfere as little as possible with the local clinical practice, where visual reading follows the FDA/EMA-approved method to assess scans. Other imaging modalities, like magnetic resonance imaging (MRI), are acquired/read according to local practice. Therefore, in this study, even if MR scans are required as part of the routine clinical workup, these may not be suitable for quantitative analysis, but only for clinical interpretation. Therefore, the use of a PET-only quantification pipeline was required.

On the other hand, the PNHS focuses entirely on measuring the natural history of amyloid accumulation in preclinical and prodromal AD populations in the context of research and disease modelling. For those purposes, accurate quantitative measurements

of Aβ burden are central to the project aims. Therefore, research-quality 3D-T1 weighted MR scans are acquired to aid the analysis of Aβ PET scans in the PNHS.

Table 1. Main characteristics of the AMYPAD trials: the Diagnostic and Patient Management Study (DPMS) and the Prognostic and Natural History Study (PNHS)

	DPMS	PNHS
EudraCT NUMBER	2017-002527-21	2018-002277-22
Recruited populations	SCD+, MCI & Dementia (as per initial syndromic diagnosis)	CU (CDR = 0) & MCIs (CDR = 0.5)
Number of sites	8	17
Number of participants	844	2000 (Target)
Longitudinal imaging	Yes (33%)	Yes (50%)
Radiotracers	[^{18}F]florbetaben and [^{18}F]flutemetamol	[^{18}F]florbetaben and [^{18}F]flutemetamol
Primary outcome	Visual Read	Centiloid
Secondary Outcomes	Centiloid, SUVR	Visual Read, SUVR, BP_{ND}
Available MRI	No (per protocol)*	Yes
Dynamic acquisitions	No (per protocol)*	Yes (up to 50%)
Data Collection	Prospective only	Prospective and retrospective

(*) Some sites have extended the general protocol locally to include MRI scans and/or dynamic PET imaging. EudraCT: European Union Drug Regulating Authorities Clinical Trials Database. SCD+: Subjective Cognitive Decline plus. MCI: Mild Cognitive Impairment. CU: Cognitively Unimpaired. CDR: Clinical Dementia Rating. SUVR: Standard Uptake Value Ratio, BP_{ND}: non-displaceable binding potential.

Quantitative comparability across Aβ tracers

To ensure adequate tracer supply and to be able to generalize results, the AMYPAD trials make use of two different commercially available Aβ PET tracers. Therefore, the comparability of Aβ burden measurements across tracers is central to achieving the goals of the two studies.

In terms of visual assessment, the approved clinical guidelines for [^{18}F]flutemetamol and [^{18}F]florbetaben are used to rate scans as positive/negative. High cross-tracer comparability of this outcome variable is expected due to similar procedures for developing the respective guidelines (*9,10*). In both cases, positivity criteria for visual reads have been validated against the presence of similar levels of pathology in autopsy samples (*11,12*). In the clinical setting, visual rating of both tracers are widely performed, even though some differences exist in the interpretation of the scans (*13*), they achieve comparable diagnostic accuracy (*14*) and high inter- and intra-rater agreement (*15*).

Quantitative Aβ measurements, on the other hand, are not directly comparable between tracers due to distinct binding and kinetic properties and thus the definition of an objective threshold for biomarker abnormality across tracers is still lacking. Among other factors, this has prevented quantitative Aβ measurements to be used in clinical practice and initiatives have been created to fill this gap in biomarker evidence (*16,17*), including the Centiloid scale (*18*) to enhance the quantitative comparability between tracers. The Centiloid method provides a standard quantitative scale to express Aβ load. In this scale, the average global uptake of any tracer in a group of healthy controls is anchored to a value of 0 Centiloid (CL) and the average of an AD-dementia group to 100 CL. To operationalize the Centiloid method, the Global Alzheimer's Association Interactive Network (GAAIN) website (http://www.gaain.org/centiloid-project) provides standard regions of interest (ROIs) for target and reference regions, a reference quantification pipeline and head-to-head reference PET scans for the different Aβ tracers. Moreover, the method includes guidance to validate new quantification pipelines using the standard PET datasets along with reference indicators for validity. Therefore, the Centiloid framework is a suitable approach to determine Aβ burden in the PNHS.

Accurate and precise Aβ quantification

Especially in longitudinal settings of early amyloid stages such as the PNHS, it is important to ensure accurate and precise quantification of Aβ burden. Unfortunately, widely used semi-quantitative methods such as the standardized uptake value ratio (SUVR) are known to be sensitive to technical and physiological confounding effects, such as scanning time window and (changes in) blood flow and clearance (*19-22*).

Full pharmacokinetic modelling can take these dependencies into account and thereby attain superior sensitivity when measuring small changes in Aβ accumulation, provided that a validated model to describe tracer kinetics is available. Fortunately, reference tissue approaches have been validated for the tracers used in AMYPAD, thus obviating the need for obtaining a metabolite corrected plasma input function (*23,24*). However, in the case of both [^{18}F]flutemetamol and [^{18}F]florbetaben, quantitative approaches typically require scanning times of up to 2 hours, resulting in a high burden for the participants and less efficient scanner and radiotracer batch usage. As an alternative, dual-phase (or dual-time window) acquisitions can be used to reduce the overall scanning time, consisting of an early (0-30 min) "dynamic" phase plus a late one that matches the timing of the standard "static" protocol: 90-110 min post-injection (p.i.). This strategy allows a break for the patient in between scanning windows and may allow for interleaved scanning

protocols (*2*), thereby optimizing tracer batch and scanner usage, whilst maintaining high quantitative accuracy. In addition, such a protocol still provides enough data to account for the effects of physiological confounders. To this end, the optimal dual-time window acquisition protocol was determined for each tracer (*25,26*) and dynamic imaging was introduced to the PNHS protocol.

In summary, the objectives and set-up of the two AMYPAD studies required the development of specific methodological approaches. These were developed such that accurate global outcome measures of Aβ load can be obtained in a comparable metric across tracers.

Retrospective harmonization of historical PET scans

One important characteristic of the PNHS is that, in addition to prospectively acquiring Aβ PET scans, it will also collect historical data from several longitudinal cohorts in Europe. When acquiring PET scans in a multicentre setting, differences in several technical factors such as scatter and attenuation correction and resolution differences of different scanners lead to heterogeneity in image properties that may introduce bias in the measured estimates of Aβ load and limit their comparability. Therefore, before pooling scans for analysis, it is important to harmonize images acquired with different scanner models to optimize the quantitative comparability of the scans acquired at the different imaging sites. This step is fundamental to maximize statistical power and to avoid site- and cohort-dependent bias.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) project has implemented and showed the viability of a method to reduce between scanner differences in multi-centre PET studies (*27*). Briefly, the method estimates the size of a Gaussian smoothing kernel that is needed to bring the effective resolution of the images acquired at each site to that of the scanner with the lowest resolution in the imaging network. The estimation of the optimal smoothing kernel for each scanner requires the acquisition of a 3D-Hoffman brain phantom. Incidentally, the use of spatial smoothing to remedy differences in PET reconstruction across sites can also be tailored to deliver a more robust analysis in addition to controlling for its impact (*28*). Adequate smoothing not only can provide better tissue delineation, but also favours the tissue delineation consistency across sites, indirectly minimising image-registration related biases.

When the PET analysis is carried out by referencing anatomical regions from a structural MRI scan, an additional source of variability stems from the lack of harmonization of the MR protocols as encountered between the parent cohorts. However, the application of a robust brain segmentation approach to MR images collected with different protocols was found to introduce an SUVR variability typically below or comparable with the scan-rescan variability (*28*). Furthermore, this variability was an order of magnitude lower than the uptake variability encountered within an amyloid-rich Region of Interest (ROI). This finding suggests that even when MR protocols differ, its impact on the PET analysis is mitigated by the actual PET resolution difference.

Aβ Quantification Methods in AMYPAD

PET acquisition protocols

Static protocol

As per the product label and established acquisition windows, static imaging protocols for both tracers were followed in the DPMS and, in a subset, also in the PNHS. These entail the acquisition of 4 frames of 5 minutes starting at 90 min. p.i. (*9,10*).

Dynamic (dual-time window) protocol

A dual-time window acquisition protocol was optimized for the tracers used in this study (*26*). In short, data from Phase II studies from each tracer were used to establish simulation parameters, and the effect of resting periods of various durations on quantitative accuracy was assessed. Although longer resting periods are generally more convenient for the patient, a window of more than 60 minutes provided (too) many outliers for both tracers and resulted in significantly larger bias in the non-displaceable binding potential (BP_{ND}), the quantitative parameter of interest. Therefore, the final and optimal dual-time window acquisition protocol for both tracers was determined to be an early scan window from 0 to 30 minutes p.i. (6x5, 3x10, 4x60, 2x150, 2x300 and 1x600 sec), followed by a 60 minutes resting period and a late scan window from 90 to 110 minutes p.i. (4x5min).

The optimized dual-time window protocol showed a maximum bias of only 3% and has the additional advantage of allowing for interleaved scanning of consecutive patients, thus facilitating efficient use of the radiotracer batches and scanner time. Furthermore, it

enables full comparability with the semi-quantitative metric by maintaining the 90-110 min p.i. window (*26*). Although most AMYPAD PNHS sites are capable and willing to perform the dual-time window protocol, the final decision for implementation is local and primarily based on logistics and capacity (*2*).

Site setup, image management and image QC

For this study, partner IXICO is responsible for scanner qualifications and data collection, hosting images on their Trial Tracker platform. This platform abides by EU/EER regulations on data protection. Images stored include both attenuation corrected and non-corrected PET for the "static" (and "dynamic" if relevant) acquisitions, as well as any CT or MR images.

Before the first subject to be scanned, a NEMA phantom is acquired to qualify the site. Aspects covered for the qualification include adherence to the specific imaging protocol, image quality issues with either the attenuation-corrected or uncorrected PET image and quantitative performance in line with published standard in the EARL accreditation guidelines (*29*).

Once sites are qualified, they perform the acquisitions on study participants and transfer the scans in DICOM format with subsequent basic metadata checks, including both the image DICOM tags and information collected in the electronic subject data forms (such as injected dose and time of injection). The metadata checks aim to ensure that data have been collected using the acquisition parameters with which the site was qualified for the study. This is an automated process that is overseen by trained analysts. Following this review, datasets are marked according to their suitability for reading and/or analysis.

For the DPMS, data marked suitable for analysis are transferred to GE. On receipt of the PET images and successful technical review, images are then subjected to a scientific quality check, where signal-to-noise ratio, reconstruction effects, subject positioning, image orientation, motion blur and structural anomalies are checked. Once the initial QC checks have been performed, data are quantified for amyloid uptake using Amypype (described below, Figure 1). Dynamic data are transferred to the Amsterdam UMC location VUmc for parametric analysis (see PNHS workflow below, Figure 2) and stored in an XNAT server.

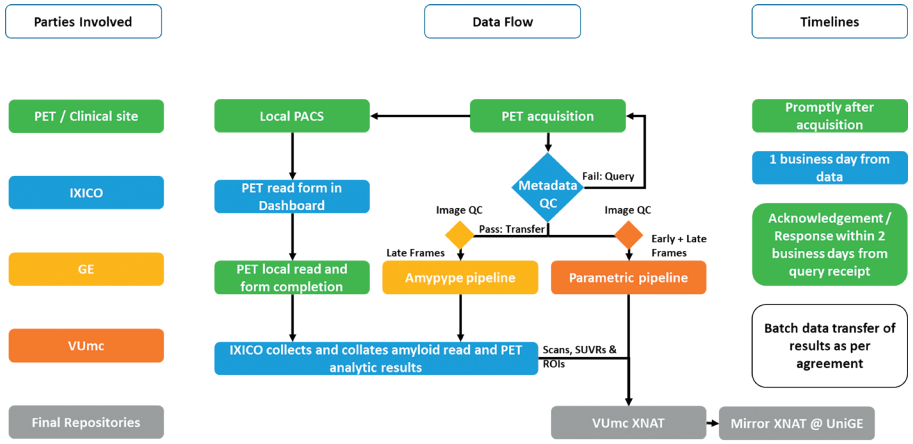


Figure 1. Imaging data flow in DPMS.

For the PNHS, data marked suitable for analysis is submitted to a visual (window) by image quality procedure includes checking that the whole brain is within the field of view, including the cerebellum, that there is no significant tilt of the head between the orbito-meatal plane and the angle made by this plane with the scanner bore axis, that the signal-to-noise ratio is acceptable - e.g. the anatomical boundaries of the brain should be clearly defined in the image, that there is no appreciable movement of the head between 5-minute frames, and that there are no brain areas with missing signal (e.g. non-perfused necrotic regions). Scans that pass this procedure are then submitted for analysis with the LEAP pipeline (described below).

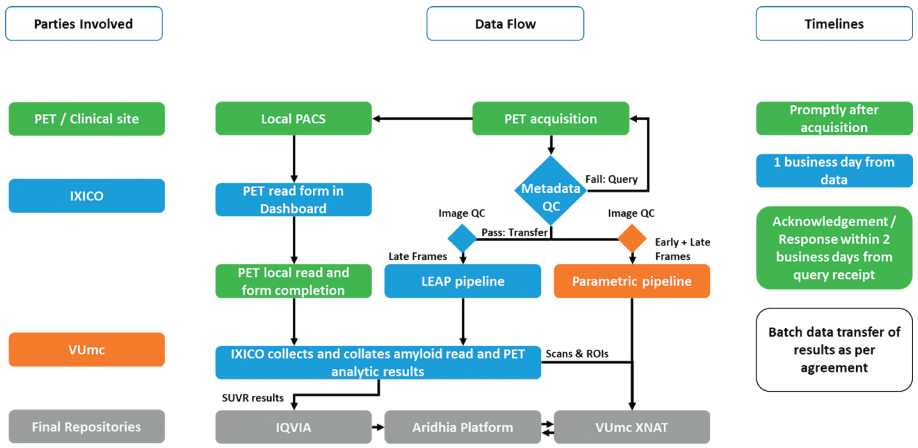


Figure 2. Imaging data flow in the PNHS.

Images acquired with the dual-phase protocol are sent to the Amsterdam UMC, location VUmc for analysis of the "dynamic" acquisition. Following receipt of the PNHS images, several quality control steps are undertaken by the VUmc image analysis team. These include a metadata check (i.e. frame start times, duration and the number of frames) and a check for possible motion between both early and late PET images and corresponding attenuation MR or low dose CT. The PET scan is then resampled into isotropic voxels, co-registered to the T1 MRI scan and the two parts of the dual-time window PET scan are combined. Next, the combined PET scan is divided into five blocks (A: frame 1-15, B: frame 16-17, C: frame 18, D: frame 19-20 and E: frame 21-22) based upon an *equal counts* principle, where each block should consist of roughly the same number of counts. The presence of patient motion between these blocks is assessed and, when necessary, corrected for. Next, the LEAP atlas, GAAIN global cortical average region (*(18)*; http://www.gaain.org/centiloid-project), and Harvard-Oxford (*30*) white matter reference region are applied to the dynamic PET scans to extract time-activity curves (TACs). Following interpolation of the reference tissue curve, the TAC is checked to confirm that both the early part (peak) and the late flat part of the curve are present and that there is a smooth interpolation during the resting period. This curve is then used together with the combined PET scan for parametric modelling purposes, and the resulting parametric images are checked visually for any abnormalities.

Harmonization of PET scans

To ensure the comparability of quantitative outcomes measured by PET in a multicentre setting, prospective harmonization is the preferred choice. It consists of setting up the reconstruction settings of each of the scanner models in the imaging network so that quantitative outcomes meet some pre-specified criteria. Examples of this strategy are widely implemented in the EANM/EARL harmonization strategies for whole-body [^{18}F]FDG imaging (*31*). At present, similar strategies are also being implemented for brain PET imaging (*32*). This is the strategy implemented with sites contributing with prospective data to the DPMS, as well as for those participating in the EPAD LCS cohort in the PNHS. However, this strategy could not be used for the whole PNHS trial as this study is also collecting historical data from existing cohorts that had already performed PET imaging using their own local, non-harmonized settings.

To overcome this limitation, an image harmonization protocol has been developed, based on the methods implemented in ADNI (*27*). Briefly, this method consists of acquiring a 3D Hoffman phantom across all of the non-EPAD LCS scanners using the reconstruction

settings applied for their historical scans. Then, different image filters can be applied to the images to match the resolution of each scanner with that of the poorest image resolution within the imaging network. The benefits of avoiding site-dependent biases outweigh the reduction of image resolution caused by the filtering process (*27*).

PET images, QC and initial quantitative results

Visual reading

In addition to the quantitative analyses described below, visual interpretation of PET images is performed using a standardized and blinded method by trained nuclear physicians or radiologists who have followed the tracer-specific reader training provided by Life Molecular Imaging (for NeuraCeq®) and/or GE HealthCare (for Vizamyl®).

For the visual read, [^{18}F]flutemetamol images are scaled to 90% of the pons signal using rainbow colour scaling, while [^{18}F]florbetaben images are scaled based on the total white matter signal and using grey colour scaling. Transverse, sagittal, and coronal views are displayed using site-specific software packages. Amyloid PET images are assessed together with an anatomical scan (i.e. T1-weighted MR or CT), if available, to limit the influence of atrophy on the visual read. The visual assessment included the frontal cortex, precuneus/posterior cingulate cortex, temporo-parietal, lateral temporal, and in the case of [^{18}F]flutemetamol the striatum.

The electronic case report form (eCRF) records the global outcome, i.e. images rated as either *positive* (VR+: unilateral binding in one or more cortical brain region or striatum) or *negative* (VR-: predominantly white matter uptake), as well as outcome per lobar region. Reader confidence of the final classification is captured on a 5-point scale (1 representing low confidence, 5 high confidence).

DPMS PET-only quantification pipeline (AmyPype)

For the DMPS study, where MRI was not routinely available, AmyPype was designed to allow the harmonized quantitative assessment of [^{18}F]flutemetamol and [^{18}F]florbetaben Aβ PET without the need for an individual's structural MRI image (*33*). AmyPype runs on a GE Advantage Workstation and uses late-phase PET-only Aβ images to output text tables of regional and composite SUVR values together with Centiloid values for both [^{18}F]flutemetamol and [^{18}F]florbetaben PET images.

In AmyPype, Aβ PET images undergo frame to frame alignment and summing, images are spatially normalized to the standard Montreal Neurological Institute (MNI152) space using an adaptive template registration method (*34*). Briefly, this method makes use of an adaptive template consisting of two tracer-specific images in the standard space, whose linear combination is capable of generating the optimal template for registering any given Aβ PET scan. An [^{18}F]florbetaben specific adaptive template has been created in AMYPAD to enable quantification with Amypype. Next, the Centiloid Cortical target template is applied to the images, as well as a grey-matter parcellated cortical-volume of interest template, consisting of the regions in the AAL atlas (*35*) which typically accumulate Aβ. The whole cerebellum is used as reference region.

Validation was achieved by comparing the output of AmyPype with the Centiloid values obtained through the GAAIN two-step validation approach (*33*). The difference between the Centiloid values obtained with AmyPype compared to the standard MR-driven pipeline was, on average, substantially less than 1% across tracers. Figure 3 provides an overview of the process. The volume of interest overlays on the PET image data is checked visually from Amypype's QC output. In case there are any observations/issues these are recorded in the quantitation report and where necessary, sites are provided feedback.

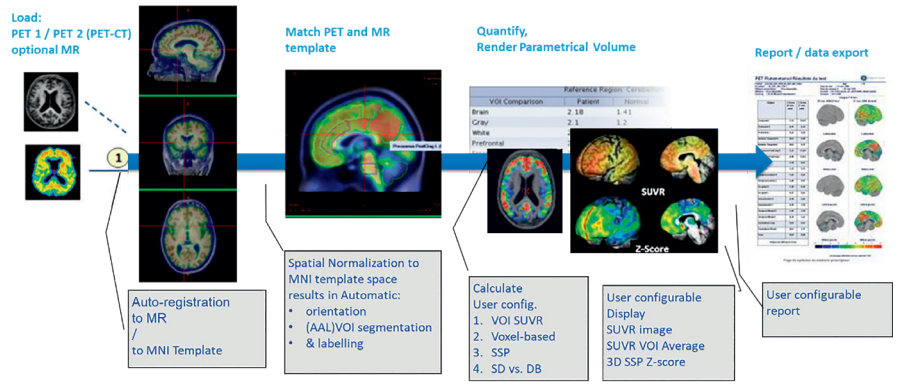


Figure 3. AmyPype: PET-only pipeline built on CortexID Suite for the analysis of the DPMS scans is based on adaptive-normalization of amyloid PET scans and outputs SUVR, Z-Scores and Centiloid values for both [^{18}F]flutememetamol and [^{18}F]florbetaben scans.

DPMS preliminary results

Figure 4 shows the frequency distribution of the Centiloid values of the first consecutive 794 scans of the DPMS. It can be observed that the data presents a bimodal distribution as would be expected from sampling of clinical populations. Colours in the bars indicate the syndromic diagnosis at the beginning of the study. A Gaussian Mixture Model (GMM) has been applied to the distribution of CL values using the Akaike Information Criterion to select the number of Gaussian functions that best model the data. For the DPMS, the optimal number of Gaussians was two, corresponding to the 'negative' (green) and 'positive' (red) groups, with expected differences by diagnostic stratum. This behaviour was observed for both tracers (data not shown).

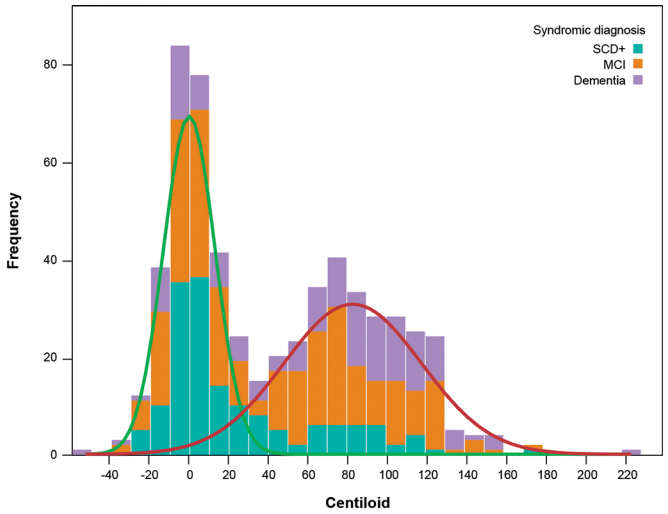


Figure 4. Preliminary distribution of Centiloid (CL) values in the first consecutive 794 participants of the DPMS.
Bars are coloured according to the baseline diagnosis of the participants (Greenish-blue: SCD+; Orange: MCI; Lilac: AD). Solid lines indicate a two Gaussian mixture model (Green: Negative group; Red: Positive Group).

PNHS MR-based quantification pipeline

For all scans marked suitable for analysis in the PNHS study, SUVR and Centiloid computation is performed using IXICO's in-house fully automated regional PET SUVR workflow. To maximize quantitative accuracy of individual Aβ load, this method employs a subject-specific multi-atlas structural MRI segmentation approach (LEAP; (*36*)) which allows for flexibly selecting multiple single and/or composite regions for the

analysis. To this end, AMYPAD leverages existing efforts in the parent cohorts to deploy a harmonized MRI acquisition protocol that will ensure comparability of MRI images acquired throughout all trial delivery centres and, hence, to maximize the precision of SUVR measurements. IXICO's PET analysis has been successfully applied to PET scans from Phase II and Phase III AD clinical trials as well as PET images from the Alzheimer's disease Neuroimaging Initiative (ADNI) and Australian Imaging, Biomarker & Lifestyle Study of Ageing (AIBL) studies. Besides, this method has been extended to fulfil the requirements of the AMYPAD-PNHS, including white matter as a reference region, and has been validated to convert SUVR to Centiloid values (*37*).

The methodology being employed consists of coregistering PET frames and creating an average image that is aligned to its corresponding MRI scan. Then, the MRI scan is parcellated using the multi-atlas LEAP methodology and these masks are used for the determination of SUVR and Centiloid metrics in native space, as shown in Figure 5. Apart from computing SUVR for regions defined in IXICO's LEAP atlas, standard space templates are incorporated to allow computation of SUVR for additional regions. In this case, standard space templates are used to define the GAAIN global cortical average region in the native space, and Harvard-Oxford white matter reference region (*30*), which is then additionally refined using morphological operations to ensure that non-white-matter tissue is excluded. The analysis is implemented in a validated and centralised computer system and overseen by IXICO in-house trained image analysts who initiate, monitor and review the component steps, as well as the final outputs of the algorithm.

SUVR is computed in the primary target regions described in Supplementary Table 1 (online) and relative to four reference regions: whole cerebellum, cerebellar grey matter, pons, and supratentorial white matter. All these regions are segmented in native space with the LEAP algorithm (i.e. not using the standard reference ROIs in GAAIN, unlike the global cortical average). White matter is included as reference region as there is ample literature supporting that supratentorial white matter, alone or in combination with some of the standard reference regions, can be more sensitive to detect longitudinal changes in Aβ burden as measured by PET (*38-46*). For each reference region, the analysis pipeline converts the regional SUVR to the appropriate tracer-specific Centiloid metrics as described in (*37*), taking into account Centiloid calibration differences due to the reference region while using the GAAIN global cortical average region as target ROI. Details on pipeline validation can be found elsewhere (*37,47*).

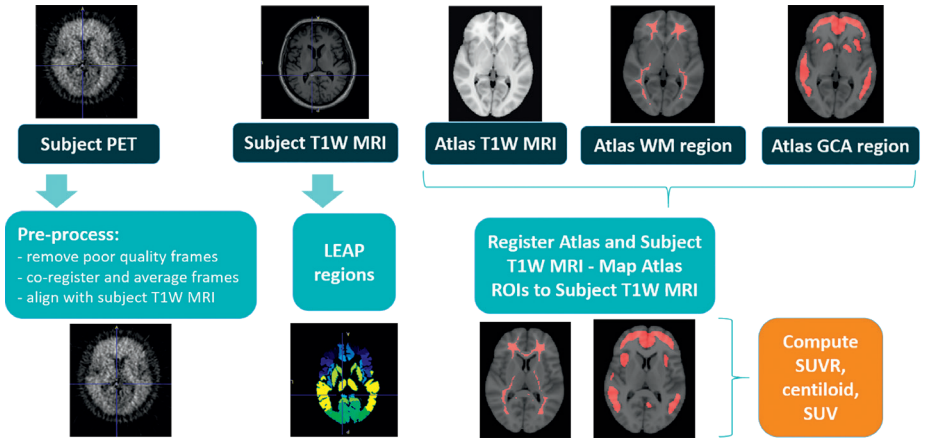


Figure 5. MR-based quantification pipeline built on LEAP for the PNHS.
This method works in the subject's space by coregistering PET scans to their corresponding MR scan which is parcellated using the LEAP method. Two target regions are calculated: the GAAIN Global Cortical Average and a composite of cortical LEAP regions. Reference regions are all based on the LEAP parcellation. This pipeline has been calibrated to render comparable Centiloid values for both [^{18}F]flutememetamol and [^{18}F]florbetaben scans, on top of SUV and SUVR values.

Since the main focus of the PNHS is to track amyloid deposition at the earliest stages, longitudinal scans are prioritized for individuals showing intermediate uptake levels. To this end, individuals are categorized into three levels: 'Negative' (if CL<12), 'Grey Zone' (if 12 ≤CL<50) and 'Positive' (if CL≥50). The lower threshold is based on recent reports of CL values optimal to exclude neuritic plaques on post-mortem samples (*48,49*). The higher cut-off value (CL=50) is based on the threshold to detect Alzheimer's disease neuropathologic change (ADNC) score of intermediate or high probability of AD (*48*).

PNHS preliminary results

The frequency of CL values of the first consecutive 248 scans of the PNHS is shown in Figure 6. The distribution for the PNHS is unimodal and skewed towards higher CL values and shows a relatively reduced range of CL values as compared to that of the DPMS. This was expected given that the PNHS is mostly recruiting cognitively unimpaired participants (CDR=0, as depicted in greenish-blue in Figure 6). A GMM was applied to the distribution of CL values using the Akaike Information Criterion to select the number of Gaussian functions that best model the data. For the PNHS, the optimal number of Gaussians was 3: one for the Negative group (green), one for the Positive group (red) and one for the intermediate or 'grey-zone' group (purple). Interestingly, the mean + 2 standard deviations of the 'Normal' Gaussian corresponds to 11.14 CL, in

line with the lower cut-off of the 'Grey Zone' category. Given the tri-modal distribution of CL values, we assume that the PNHS is successfully recruiting a sample enriched for individuals displaying incipient amyloid pathology.

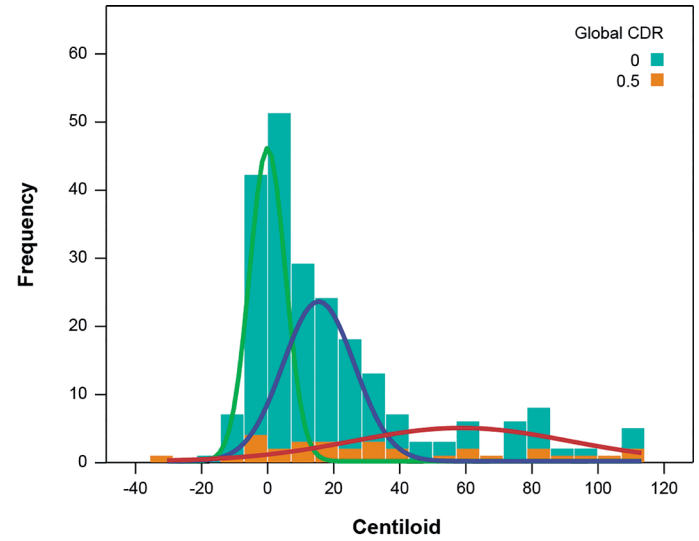


Figure 6. Preliminary distribution of CL values in the first consecutive 248 of the PNHS.
Bars are stacked by the Clinical Dementia Rating (CDR) status of the participants (Greenish-blue: CDR=0; Orange: CDR=0.5). Solid lines indicate a three-Gaussian Mixture Model fitted on the frequency data. Akaike information criterion showed that 3 was the optimal number of Gaussians for this dataset. The Green line corresponds to the Negative group; the Purple one to the intermediate of 'Grey Zone' and the Red one to the positive group. These results show that the PNHS is successful in recruiting a very high percentage of individuals with intermediate amyloid levels.

Dynamic PET quantification pipeline

The Amsterdam UMC location VUmc, is responsible for obtaining fully quantitative estimates of Aβ load for each individual scanned according to the dual-time window protocol within the AMYPAD studies. To allow for direct comparability with the static quantification pipeline, results from IXICO's pre-processing pipeline are incorporated into the quantitative pipeline for dynamic scans.

To this end, IXICO transfers the dual-time window (0-30 and 90-110 min p.i.) PET images (NIfTI and DICOM format), corresponding T1 MR images and subject-specific templates (all NIfTI format): the LEAP atlas, GAAIN global cortical average region,

and Harvard-Oxford white matter reference region Then, from the DICOM PET images, information related to frame times, frame prompts and image rescale factors are extracted. Next, PET images (NIfTI) are resampled into isotropic voxels (2x2x2 mm), co-registered to the T1 MRI scan and the two parts of the dual-time window PET scan are combined. In the case of between-frame motion, corrections are applied using software developed in-house. Next, the LEAP template image is applied to the resulting PET scan to obtain a reference TAC of the cerebellar grey matter for kinetic modelling.

Missing data points of the reference tissue TAC are interpolated using the 2T4k_V_b model together with a typical, tracer-specific, input function (*26*). In the case of sub-optimal interpolation, cubic interpolation is used as an alternative. Voxel-based pharmacokinetic analyses are then carried out using the PPET software tool (*50*) using the SRTM2 model to derive estimates of both the BP_{ND} and a measure of relative perfusion (R_1). Next, subject-specific templates are applied to obtain global as well as regional BP_{ND} and R_1 values.

In order to enable quantification of full dynamic or dual-phase acquisitions, in AMYPAD a dedicated open-source software suite is being developed. It will enable MR-assisted and PET-only pre-processing pipelines, kinetic modelling using SRTM models, and Centiloid calibration.

Discussion

The two AMYPAD trials, the DPMS and the PNHS, are expected to provide significant contributions to extend the use and value of amyloid PET both in clinical and research settings. To achieve their respective goals, these two trials are acquiring cross-sectional and longitudinal PET data with either [^{18}F]florbetaben or [^{18}F]flutemetamol in thousands of participants across the AD clinical severity *continuum*, with an emphasis on preclinical and prodromal participants, but also including dementia stages. By doing so, this project is uniquely placed to develop quantitative methodologies to measure Aβ load by PET as well as to determine the optimal quantitative approach for a wide range of applications in clinical and research settings. The main opportunities for AMYPAD to fill existing gaps in the body of knowledge on quantitative Aβ imaging pivot around: i) the assessment of different measures of Aβ and ii) the capacity of dynamic PET imaging to provide improved performance and additional data as compared with static imaging.

Comparability of Aβ measures

One limiting factor for multicentre Aβ PET studies lies in the geographical availability of radiotracers to cover all clinical sites. All quantitative estimates of Aβ burden in AMYPAD are expressed in CL units to ensure quantitative comparability between the two tracers in this study as well as with data acquired by other studies. The comparability of CL units obtained with the two tracers will be verified by different methods, including the assessment of agreement of CL cut-offs against visual reads.

While the CL framework has been designed to account for the confounding effects of different tracers or processing pipelines in quantification, it does not explicitly correct for the impact that variation of spatial resolution between different scanners may have on quantification. This must be addressed either prospectively by harmonizing reconstruction settings or retrospectively by applying post-processing algorithms. In this respect, AMYPAD has established a collaboration agreement with EARL (EANM Research Ltd), an initiative of the European Association of Nuclear Medicine. At the moment, EARL provides several accreditation programs for nuclear medicine imaging including procedures for standardization and harmonization of whole-body [^{18}F]FDG PET imaging (*29*). This accreditation ensures the quantitative consistency of SUV measurements in multicentre studies. Thanks to the partnership between AMYPAD and EARL, AMYPAD scans will be used to validate the new brain PET accreditation program that EARL is developing. At the same time, amyloid PET measurements in AMYPAD will be accredited by the reference specifications stipulated by EARL.

As a clinical biomarker, it may, in certain situations, be necessary to dichotomize the continuous measure of amyloid burden into negative and positive (A-/+) categories (*51*). Thanks to the Centiloid framework providing a reference scale to measure Aβ load, cut-off values to dichotomize participants into A-/+ can be generalized across studies. Following this rationale, cut-offs have now been derived to maximize the agreement between Aβ PET and several other measurements (or criteria) to detect the presence of Aβ pathology. Reference CL cut-offs have been derived against neuropathology (*48,49*), visual read positivity (*52*), Aβ abnormality as determined in Cerebrospinal Fluid (CSF) (*53*) or to predict future Aβ accumulation (*51*). Interestingly, all these cut-offs seem to gather around relatively low CL values between 10-20 CL. However, these quantitative cut-offs need to be taken with caution since it has been reported that Centiloid conversion might not be able to eliminate all differences in amyloid burden measurements due to variabilities in underlying acquisition protocols and analysis techniques (*54*). In this regard, some improvements to

the Centiloid method and alternative methods to enable the comparability of quantitative Aβ measurements across tracers have also been developed *(55-57)*.

In the clinical setting, the quantification of Aβ PET has been suggested to improve the confidence of visual reads (*58*). In AMYPAD, the confidence of visual reads is also recorded and will be analysed in association with quantitative outcomes. We expect that confidence will be lower for scans with CL close to the cut-offs for positivity against visual reads. The value of quantification to improve the accuracy and confidence of visual assessment of amyloid PET scans remains an open question.

When dichotomizing subjects into A-/+ status with more than one method, a prolific line of research has been the study of discordant cases (*59-63*). AMYPAD will enable the study of A-/+ discrepant cases as defined by visual inspection and CL cut-offs or between PET and CSF, which is available in some parent cohorts. Since the agreement between these Aβ categorization approaches is very high, a large number of participants need to be scanned to capture a sufficiently high number of discordant cases to enable the study of the technical, demographic and/or pathophysiological factors associated with such a mismatch. The high number of scans acquired in AMYPAD may contribute to this matter. Another unique aspect of AMYPAD is the focus on preclinical and prodromal PNHS populations in which the percentage of discordant cases is expected to be higher than in case-control studies of patients with overt AD dementia.

Reference cut-offs can be used to operationalize inclusion or follow-up of studies and trials with different goals. For example, since the PNHS has a strong focus on the earliest stages of Aβ accumulation, follow-up scans will preferentially be performed in individuals between 12 and 50 CL. This approach is already revealing three ordinal categories of Aβ accumulation, i.e. negative, grey-zone and positive, thus extending the typical binary classifications. For example, we show that Aβ accumulation can be staged with more than two categories (i.e. A 0/1/2 rather than A-/+) similar to the TNM staging system in oncology (*64*). It is worth noting that such an approach was already contemplated in the ATN framework (*65*). In the same vein, a topographical staging methodology has been developed in AMYPAD and successfully applied to 4722 external PET scans acquired with four different tracers (*66*). This method allowed the staging of over 99% of the scans in the validation set and it has been shown that increasing stage was associated with faster cognitive decline. This staging method will be applied to several research questions in AMYPAD to explore in what cases it may prove to bring value over

global Aβ measurements. Also, a refinement of this method is underway by defining the topography of the stages at the voxel level, thus circumventing the dependency on *a priori* parcellation with a brain atlas (*67*). Notably, this strategy has also been extended to visual reading of the scans and a 3-scale model has been developed in AMYPAD by grading regional positivity in the scans (*52*). Future studies will focus on the prognostic value of these grading schemes.

Dynamic Aβ PET imaging

As efforts to tackle AD shift towards prevention, a better understanding of the pathophysiological events within preclinical stages and their downstream consequences is needed. This calls for novel methodological and analytical approaches that maximize sensitivity to early changes. The AMYPAD PNHS focuses on studying the early stages of Aβ accumulation in preclinical and prodromal populations. From a methodological perspective, one of the most salient characteristics of this study is the acquisition of a considerable amount of dynamic PET scans in a multi-centre study. By accounting for several technical and physiological factors, dynamic PET quantitative measures (DVR or BP_{ND}) typically show reduced variability and bias as compared with simplified (semi-)quantitative ones such as SUVR (*68*). These improvements result in higher statistical power for a given sample size or, equivalently, reduced sample sizes are needed to detect a given effect size. For example, it has been shown that dynamic amyloid PET imaging may improve statistical power in prevention clinical trials in individuals with early amyloid accumulation (*69*).

However, the utilization of dynamic PET imaging in large multi-centre trials has typically been considered to be disadvantageous from a cost-benefit perspective. Indeed, compared with static acquisitions, dynamic PET is more costly due to lower efficiency in the use of radiotracer batches and longer acquisition times, which also increases participant burden. To alleviate these limitations, a dual-time window acquisition protocol has been developed in AMYPAD (*26*). This acquisition protocol has been successfully deployed in 5 sites across Europe using both [^{18}F]flutemetamol and [^{18}F]florbetaben in a substantial number of cases. Thanks to the acquisition of such a large multi-centre dynamic PET dataset, AMYPAD will be able to determine the cost-benefit of dynamic PET scanning for different relevant clinical trial scenarios and detect those cases in which it may deliver improved performance as compared to the static protocol.

It is worth noting that reference dynamic [^{11}C]PiB PET scans are provided on the GAAIN website. Indeed, one of the examples in the original Centiloid paper consisted

of the calibration of BP_{ND} data against standard SUVR [^{11}C]PiB units (*18*). As compared to similar SUVR-based CL pipelines with [^{11}C]PiB data, a BP_{ND}-based CL metric has been reported to show reduced variability while keeping a very high effect size to detect longitudinal changes in Aβ deposition (*54*). In this respect, AMYPAD will investigate optimal methods to extend the CL framework to dynamic scans acquired with [^{18}F]florbetaben and [^{18}F]flutemetamol.

An additional benefit of dynamic imaging is that it provides surrogate markers of cerebral blood flow (CBF). Maps of the relative tracer delivery rate (R_1) of Aβ tracers or their early frames have shown to be valid surrogate markers of relative CBF (*70,71*). CBF is a relevant biomarker for AD since reduced CBF has consistently been found in brain regions typically associated with lower glucose metabolism and atrophy in AD patients (*72*). Also, unlike elevated levels of cerebral Aβ, CBF reductions have been found to correlate with cognitive decline in AD patients (*73,74*) and they already occur in preclinical AD stages (*22,73,74*). Taken together, these findings support that CBF maps can be used in clinical trials to recruit cognitively unimpaired individuals which may undergo cognitive decline shortly. Another potential use of these perfusion maps may be the differential diagnosis of dementia syndromes in combination with measurements of Aβ burden. Relevant to these applications, dynamic Aβ PET imaging can derive information on both relative CBF and amyloid pathology from a single scanning session.

Conclusions

We report the challenges and solutions adopted for acquisition, management and analysis of Aβ PET in the two multicentre trials within AMYPAD and present early results pointing out that the two trials are successful in recruiting their respective target populations. Results of these trials will address the utility of Aβ PET in clinical and research settings. Additionally, AMYPAD will provide methodological contributions to improve accuracy, precision and comparability of Aβ load measured by PET. The trials have adopted the Centiloid framework to render comparable estimates of Aβ load, irrespective of the radiotracer used. In addition, the PNHS will be able to assess whether, and for which purposes, dynamic imaging can be advantageous to methods based on the standard static acquisition protocol. These developments have the overarching goal of achieving more sensitive and reproducible measurements of Aβ burden and their longitudinal change as measured by Aβ PET.

References

1. Frisoni GB, Barkhof F, Altomare D, Berkhof J, Boccardi M, Canzoneri E, et al. AMYPAD Diagnostic and Patient Management Study: Rationale and design. Alzheimer's & dementia: the journal of the Alzheimer's Association. 2019;15:388-99.
2. Lopes Alves I, Collij LE, Altomare D, Frisoni GB, Saint-Aubert L, Payoux P, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2020;16:750-8.
3. Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW, Consortium E. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. BMJ Open. 2019;8:e021017.
4. Konijnenberg E, Carter SF, Ten Kate M, den Braber A, Tomassen J, Amadi C, et al. The EMIF-AD PreclinAD study: study design and baseline cohort overview. Alzheimer's research & therapy. 2018;10:75.
5. Moreno-Grau S, Rodriguez-Gomez O, Sanabria A, Perez-Cordon A, Sanchez-Ruiz D, Abdelnour C, et al. Exploring APOE genotype effects on Alzheimer's disease risk and amyloid beta burden in individuals with subjective cognitive decline: The FundacioACE Healthy Brain Initiative (FACEHBI) study baseline results. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2018;14:634-43.
6. Molinuevo JL, Gramunt N, Gispert JD, Fauria K, Esteller M, Minguillon C, et al. The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease. Alzheimers Dement (N Y). 2016;2:82-92.
7. Adamczuk K, De Weer AS, Nelissen N, Dupont P, Sunaert S, Bettens K, et al. Functional Changes in the Language Network in Response to Increased Amyloid beta Deposition in Cognitively Intact Older Adults. Cereb Cortex. 2017;27:3879.
8. Hanseeuw BJ, Malotaux V, Dricot L, Quenon L, Sznajer Y, Cerman J, et al. Defining a Centiloid scale threshold predicting long-term progression to dementia in patients attending the memory clinic: an [(18)F] flutemetamol amyloid PET study. European journal of nuclear medicine and molecular imaging. 2021;48:302-10.
9. Life Radiopharma Berlin GmbH. Neuraceq : EPAR - Product Information. Summary of product characteristics. EMEA/H/C/002553 - IB/0027/G2019.
10. GE Healthcare AS. Vizamyl : EPAR - Product Information. Summary of product characteristics. EMEA/H/C/002557 - II/0022/G2020.
11. Ikonomovic MD, Buckley CJ, Heurling K, Sherwin P, Jones PA, Zanette M, et al. Post-mortem histopathology underlying beta-amyloid PET imaging following flutemetamol F 18 injection. Acta Neuropathol Commun. 2016;4:130.

12. Sabri O, Sabbagh MN, Seibyl J, Barthel H, Akatsu H, Ouchi Y, et al. Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: phase 3 study. Alzheimer's & dementia: the journal of the Alzheimer's Association. 2015;11:964-74.
13. Lundeen TF, Seibyl JP, Covington MF, Eshghi N, Kuo PH. Signs and Artifacts in Amyloid PET. Radiographics. 2018;38:2123-33.
14. Morris E, Chalkidou A, Hammers A, Peacock J, Summers J, Keevil S. Diagnostic accuracy of (18)F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. European journal of nuclear medicine and molecular imaging. 2016;43:374-85.
15. Paghera B, Altomare D, Peli A, Morbelli S, Buschiazzo A, Bauckneht M, et al. Comparison of visual criteria for amyloid-PET reading: could their merging reduce the inter-raters variability? Q J Nucl Med Mol Imaging. 2019.
16. Chiotis K, Saint-Aubert L, Boccardi M, Gietl A, Picco A, Varrone A, et al. Clinical validity of increased cortical uptake of amyloid ligands on PET as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. Neurobiol Aging. 2017;52:214-27.
17. Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. Lancet Neurol. 2017;16:661-76.
18. Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD, Sr., Jagust WJ, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2015;11:1-15 e1-4.
19. van Berckel BN, Ossenkoppele R, Tolboom N, Yaqub M, Foster-Dingley JC, Windhorst AD, et al. Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2013;54:1570-6.
20. Cselenyi Z, Farde L. Quantification of blood flow-dependent component in estimates of beta-amyloid load obtained using quasi-steady-state standardized uptake value ratio. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2015;35:1485-93.
21. Heeman F, Yaqub M, Lopes Alves I, Heurling K, Bullich S, Gispert JD, et al. Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2020:271678X20918029.
22. Binnewijzend MA, Benedictus MR, Kuijer JP, van der Flier WM, Teunissen CE, Prins ND, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. Eur Radiol. 2016;26:506-14.

23. Becker GA, Ichise M, Barthel H, Luthardt J, Patt M, Seese A, et al. PET quantification of 18F-florbetaben binding to beta-amyloid deposits in human brains. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2013;54:723-31.
24. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [18F]flutemetamol. NeuroImage. 2015;121:184-92.
25. Bullich S, Barthel H, Koglin N, Becker GA, De Santi S, Jovalekic A, et al. Validation of Noninvasive Tracer Kinetic Analysis of [18F]-Florbetaben PET Using a Dual-Time-Window Acquisition Protocol. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2018;59:1104-10.
26. Heeman F, Yaqub M, Lopes Alves I, Heurling K, Berkhof J, Gispert JD, et al. Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. EJNMMI research. 2019;9:32.
27. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. NeuroImage. 2009;46:154-9.
28. Palombit A, Manber R, Joules R, Wolz R. Amyloid PET variability due to variation in MRI protocol and anatomical segmentation. Human Amyloid Imaging (HAI). Miami (USA)2020.
29. Kaalep A, Sera T, Oyen W, Krause BJ, Chiti A, Liu Y, et al. EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems. European journal of nuclear medicine and molecular imaging. 2018;45:412-22.
30. Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 2006;31:968-80.
31. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. European journal of nuclear medicine and molecular imaging. 2017;44:17-31.
32. Verwer EE, Golla SVS, Kaalep A, Lubberink M, van Velden FHP, Bettinardi V, et al. Harmonisation of PET/CT Performance for Brain Studies. Annual Congress of the European Association of Nuclear Medicine. Barcelona2019.
33. Buckley CJ, Foley C, Battle M, Grecchi E, Farrar G, Gispert JD, et al. AmyPype: an automated system to quantify AMYPAD's [18F]flutemetamol and [18F]florbetaben images including regional SUVR and Centiloid analysis. Congress of the European Association of Nuclear Medicine (EANM) Barcelona (Spain)2019.
34. Lundqvist R, Lilja J, Thomas BA, Lotjonen J, Villemagne VL, Rowe CC, et al. Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2013;54:1472-8.

35. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. NeuroImage. 2020;206:116189.
36. Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D, Alzheimer's Disease Neuroimaging I. LEAP: learning embeddings for atlas propagation. NeuroImage. 2010;49:1316-25.
37. Grecchi E, Foley C, Gispert JD, Wolz R. CENTILOID PET SUVR ANALYSIS USING THE SUPRATENTORIAL WHITE MATTER AS REFERENCE REGION. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2018;14:P1278.
38. Blautzik J, Brendel M, Sauerbeck J, Kotz S, Scheiwein F, Bartenstein P, et al. Reference region selection and the association between the rate of amyloid accumulation over time and the baseline amyloid burden. European journal of nuclear medicine and molecular imaging. 2017;44:1364-74.
39. Brendel M, Hogenauer M, Delker A, Sauerbeck J, Bartenstein P, Seibyl J, et al. Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. NeuroImage. 2015;108:450-9.
40. Chen K, Roontiva A, Thiyyagura P, Lee W, Liu X, Ayutyanont N, et al. Improved power for characterizing longitudinal amyloid-beta PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2015;56:560-6.
41. Fleisher AS, Joshi AD, Sundell KL, Chen YF, Kollack-Walker S, Lu M, et al. Use of white matter reference regions for detection of change in florbetapir positron emission tomography from completed phase 3 solanezumab trials. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2017;13:1117-24.
42. Landau SM, Fero A, Baker SL, Koeppe R, Mintun M, Chen K, et al. Measurement of longitudinal beta-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2015;56:567-74.
43. Schwarz CG, Senjem ML, Gunter JL, Tosakulwong N, Weigand SD, Kemp BJ, et al. Optimizing PiB-PET SUVR change-over-time measurement by a large-scale analysis of longitudinal reliability, plausibility, separability, and correlation with MMSE. NeuroImage. 2017;144:113-27.
44. Su Y, Blazey TM, Owen CJ, Christensen JJ, Friedrichsen K, Joseph-Mathurin N, et al. Quantitative Amyloid Imaging in Autosomal Dominant Alzheimer's Disease: Results from the DIAN Study Group. PloS one. 2016;11:e0152082.
45. Tryputsen V, DiBernardo A, Samtani M, Novak GP, Narayan VA, Raghavan N, et al. Optimizing regions-of-interest composites for capturing treatment effects on brain amyloid in clinical trials. Journal of Alzheimer's disease : JAD. 2015;43:809-21.

46. Salvadó G, Foley C, Grecchi E, Cardoso JM, Lopes Alves I, Markiewicz P, et al. Supratentorial white matter appears to be a better reference region than standard infratentorial regions for longitudinal quantification of [18F]flutemetamol scans. Clinical Trials in Alzheimer's Disease (CTAD). Barcelona (Spain)2018.
47. Grecchi E, Wolz R, Hill D. Operationalizing Automated PET SUVR Quantification in Multi-Centre AD Clinical Trials. Alzheimer's Association International Conference (AAIC). London (UK)2017.
48. Amadoru S, Dore V, McLean CA, Hinton F, Shepherd CE, Halliday GM, et al. Comparison of amyloid PET measured in Centiloid units with neuropathological findings in Alzheimer's disease. Alzheimer's research & therapy. 2020;12:22.
49. La Joie R, Ayakta N, Seeley WW, Borys E, Boxer AL, DeCarli C, et al. Multisite study of the relationships between antemortem [(11)C]PiB-PET Centiloid values and postmortem measures of Alzheimer's disease neuropathology. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2019;15:205-16.
50. Bos I, Vos S, Vandenberghe R, Scheltens P, Engelborghs S, Frisoni G, et al. The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. Alzheimer's research & therapy. 2018;10:64.
51. Jack CR, Jr., Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2017;13:205-16.
52. Collij L, Salvadó G, Lopes Alves I, Reimand J, Wink AM, Zwan M, et al. Examining Centiloid Quantification against Visual Assessment using [18 F]flutemetamol PET. Alzheimer's Association International Conference. Amsterdam, The Netherlands2020.
53. Salvado G, Molinuevo JL, Brugulat-Serrat A, Falcon C, Grau-Rivera O, Suarez-Calvet M, et al. Centiloid cut-off values for optimal agreement between PET and CSF core AD biomarkers. Alzheimer's research & therapy. 2019;11:27.
54. Su Y, Flores S, Hornbeck RC, Speidel B, Vlassenko AG, Gordon BA, et al. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. Neuroimage Clin. 2018;19:406-16.
55. Chincarini A, Peira E, Morbelli S, Pardini M, Bauckneht M, Arbizu J, et al. Semi-quantification and grading of amyloid PET: A project of the European Alzheimer's Disease Consortium (EADC). Neuroimage Clin. 2019;23:101846.
56. Cho SH, Choe YS, Kim HJ, Jang H, Kim Y, Kim SE, et al. Correction to: A new Centiloid method for (18)F-florbetaben and (18)F-flutemetamol PET without conversion to PiB. European journal of nuclear medicine and molecular imaging. 2020;47:1610.

57. Properzi MJ, Buckley RF, Chhatwal JP, Donohue MC, Lois C, Mormino EC, et al. Nonlinear Distributional Mapping (NoDiM) for harmonization across amyloid-PET radiotracers. NeuroImage. 2019;186:446-54.
58. Pontecorvo MJ, Arora AK, Devine M, Lu M, Galante N, Siderowf A, et al. Quantitation of PET signal as an adjunct to visual interpretation of florbetapir imaging. European journal of nuclear medicine and molecular imaging. 2017;44:825-37.
59. Doecke JD, Ward L, Burnham SC, Villemagne VL, Li QX, Collins S, et al. Elecsys CSF biomarker immunoassays demonstrate concordance with amyloid-PET imaging. Alzheimer's research & therapy. 2020;12:36.
60 Landau SM, Lu M, Joshi AD, Pontecorvo M, Mintun MA, Trojanowski JQ, et al. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of beta-amyloid. Ann Neurol. 2013;74:826-36.
61. Leuzy A, Chiotis K, Hasselbalch SG, Rinne JO, de Mendonca A, Otto M, et al. Pittsburgh compound B imaging and cerebrospinal fluid amyloid-beta in a multicentre European memory clinic study. Brain. 2016;139:2540-53.
62. Mattsson N, Insel PS, Donohue M, Landau S, Jagust WJ, Shaw LM, et al. Independent information from cerebrospinal fluid amyloid-beta and florbetapir imaging in Alzheimer's disease. Brain. 2015;138:772-83.
63. Palmqvist S, Zetterberg H, Mattsson N, Johansson P, Alzheimer's Disease Neuroimaging I, Minthon L, et al. Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. Neurology. 2015;85:1240-9.
64. Brierley J, Gospodarowicz MK, Wittekind C. TNM classification of malignant tumours. Eighth edition. ed. Chichester, West Sussex, UK ; Hoboken, NJ: John Wiley & Sons, Inc.; 2017.
65. Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology. 2016;87:539-47.
66. Collij LE, Heeman F, Salvado G, Ingala S, Altomare D, de Wilde A, et al. Multitracer model for staging cortical amyloid deposition using PET imaging. Neurology. 2020;95:e1538-e53.
67. Salvadó G, Collij L, Ninerola-Baizan A, Perissinotti A, Van der Flier WM, Visser PJ, et al. Voxel-Based Amyloid PET Staging for the Whole Alzheimer's Disease Continuum. Alzheimer's Association International Conference (AAIC). Los Angeles (USA)2019.
68. Lopresti BJ, Klunk WE, Mathis CA, Hoge JA, Ziolko SK, Lu X, et al. Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2005;46:1959-72.
69. Lopes Alves I, Heeman F, Collij LE, et al. Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging. *Alzheimers Res Ther*. 2021;13:82

70. Bilgel M, Beason-Held L, An Y, Zhou Y, Wong DF, Resnick SM. Longitudinal evaluation of surrogates of regional cerebral blood flow computed from dynamic amyloid PET imaging. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2020;40:288-97.
71. Ottoy J, Verhaeghe J, Niemantsverdriet E, De Roeck E, Wyffels L, Ceyssens S, et al. (18) F-FDG PET, the early phases and the delivery rate of (18)F-AV45 PET as proxies of cerebral blood flow in Alzheimer's disease: Validation against (15)O-H2O PET. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2019;15:1172-82.
72. Matsuda H. Cerebral blood flow and metabolic abnormalities in Alzheimer's disease. Ann Nucl Med. 2001;15:85-92.
73. Benedictus MR, Leeuwis AE, Binnewijzend MA, Kuijer JP, Scheltens P, Barkhof F, et al. Lower cerebral blood flow is associated with faster cognitive decline in Alzheimer's disease. Eur Radiol. 2017;27:1169-75.
74. Leeuwis AE, Benedictus MR, Kuijer JPA, Binnewijzend MAA, Hooghiemstra AM, Verfaillie SCJ, et al. Lower cerebral blood flow is associated with impairment in multiple cognitive domains in Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2017;13:531-40.

3.

Optimized dual-time-window protocols for quantitative [^{18}F]flutemetamol and [^{18}F]florbetaben PET studies

Fiona Heeman, Maqsood Yaqub, Isadora Lopes Alves, Kerstin Heurling, Johannes Berkhof, Juan Domingo Gispert, Santiago Bullich, Christopher Foley, Adriaan A. Lammertsma
On behalf of the AMYPAD Consortium

***As published in EJNNMI Research*, 2019; 9:32**

Abstract

Background: A long dynamic scanning protocol may be required to accurately measure longitudinal changes in amyloid load. However, such a protocol results in a lower patient comfort and scanning efficiency compared with static scans. A compromise can be achieved by implementing dual-time window protocols. This study aimed to optimize these protocols for quantitative [^{18}F]flutemetamol and [^{18}F]florbetaben studies.

Methods: Rate constants for subjects across the Alzheimer's disease spectrum (i.e. non-displaceable binding potential (BP_{ND}) in the range 0.02-0.77 and 0.02-1.04 for [^{18}F]flutemetamol and [^{18}F]florbetaben, respectively) were established based on clinical [^{18}F]flutemetamol (N=6) and [^{18}F]florbetaben (N=20) data, and used to simulate tissue time-activity curves (TACs) of 110 minutes using a reference tissue and plasma input model. Next, noise was added (N=50) and data points corresponding to different intervals were removed from the TACs, ranging from 0 (i.e. 90-90 = full-kinetic curve) to 80 (i.e. 10-90) minutes, creating a dual-time window. Resulting TACs were fitted using the simplified reference tissue method (SRTM) to estimate the BP_{ND}, outliers ($\geq 1.5 * BP_{ND}$ max) were removed and the bias was assessed using the distribution volume ratio (DVR=BP_{ND}+1). To this end, acceptability curves, which display the fraction of data below a certain bias threshold, were generated and the area under those curves were calculated.

Results: [^{18}F]Flutemetamol and [^{18}F]florbetaben data demonstrated an increased bias in amyloid estimate for larger intervals and higher noise levels. An acceptable bias (≤3.1%) in DVR could be obtained with all except the 10-90 and 20-90 minutes intervals. Furthermore, a reduced fraction of acceptable data and most outliers were present for these two largest intervals (maximum percentage outliers 48 and 32 for [^{18}F]flutemetamol and [^{18}F]florbetaben, respectively).

Conclusions: The length of the interval inversely correlates with the accuracy of the BP_{ND} estimates. Consequently, a dual-time window protocol of 0-30 and 90-110 minutes (= maximum of 60 minutes interval) allows for accurate estimation of BP_{ND} values for both tracers.

Keywords: Amyloid, quantification, simplified methods, flutemetamol PET, florbetaben PET, Alzheimer's disease

Background

Deposition of amyloid-beta (Aβ) plaques in the brain is the earliest *in vivo* measurable hallmark in the development of Alzheimer's disease (AD), which is the most common type of dementia (*1–3*). Therefore, visualisation of Aβ deposits *in vivo* is essential for improving early diagnosis and monitoring treatment effects (*4*). To this end, various positron emission tomography (PET) amyloid tracers have been developed (*5*). Among those, fluorine-18 (18F) labelled tracers approved by the European Medicines Agency (EMA)/Food and Drug Administration (FDA) are of special interest for clinical trials due to their relatively long half-life $t_{1/2}$ = 109.8min compared to [^{11}C]PiB (Carbon-11 Pittsburgh Compound B) and commercial availability (*5,6*).

In addition to visualization, amyloid PET allows for quantification of underlying physiological processes, such as the level of Aβ plaque burden (*7–10*). For diagnostic purposes, a static scan acquired at pseudo-equilibrium, using a tracer specific approved method, has been deemed sufficient in combination with visual assessment of the images. In research settings, this simplified protocol is commonly used to calculate the standardized uptake value ratio (SUVR) (*5*). SUVR, however, is only a semi-quantitative parameter that is known to be affected by both scanning time window and (changes in) blood flow (*11,12*). Given this dependency, full quantification using pharmacokinetic modelling may be required to obtain higher overall sensitivity for measuring longitudinal changes (e.g. for monitoring disease progression or treatment response), especially during the early stages of the disease when amyloid is still accumulating. Pharmacokinetic modelling, however, requires a dynamic scanning protocol, which can last for up to two hours depending on the actual tracer. These long acquisition protocols result in lower patient comfort and less efficient use of both scanner and tracer batch, in addition to an increased risk of motion artifacts. Dynamic data acquisition in a dual-time window protocol (also called "coffee-break" protocol), however, can be used to reduce overall scanning time, in which data are acquired separately for early and late phases. Such a protocol provides a resting period for the patient and, when long enough, may also allow for interleaved scanning protocols, thereby optimizing tracer batch and scanner usage (i.e. costs), while maintaining a high quantitative accuracy.

So far, some studies have used a dual-time window protocol using static acquisitions of amyloid-PET data. An early scan (i.e. 0-10 minutes p.i.) was proposed in addition to the (standard) late static scan, as it has been reported that the early scan may provide

information on metabolism and neuronal injury, possibly circumventing the need for additional [^{18}F]FDG imaging (*12–16*). Recently, Bullich and colleagues demonstrated that the non-displaceable binding potential (BP_{ND}) obtained using a dual-time window acquisition protocol (0-30 and 120-140 minutes p.i.) correlated well with BP_{ND} obtained using a full dynamic acquisition protocol of 140 minutes (*17*). This [^{18}F]florbetaben study, however, did not report details about different resting periods, nor did it assess the robustness of the dual-time window protocol for subjects across the AD spectrum and for different noise levels (i.e. for regions of different sizes).

The purpose of the present simulation study was to define optimal dual-time window acquisition protocols for [^{18}F]florbetaben and [^{18}F]flutemetamol, both in terms of patient comfort and throughput, whilst maintaining high quantitative accuracy. These simulations were focused on early stages of the disease, given the potential value of amyloid imaging to guide interventions aimed at secondary prevention of AD dementia.

Methods

Subjects and PET data

[^{18}F]Flutemetamol whole blood input curves, metabolite corrected arterial plasma input curves and time-activity curves (TACs) from 12 volumes of interest (VOIs) of three healthy controls and three probable AD subjects were obtained from Heurling et al. and Nelissen et al. (*7,18*). [^{18}F]Florbetaben metabolite corrected and uncorrected plasma input curves together with whole blood samples and TAC data from 13 VOIs of ten healthy controls and ten AD subjects were obtained from Becker et al. (*8*).

Kinetic models for BP_{ND} estimation

It has been shown that the reversible two tissue compartment model (4 rate constants) with additional blood volume fraction parameter (2T4k_V_b) is the optimal plasma input model for describing both [^{18}F]flutemetamol and [^{18}F]florbetaben kinetics (*8,18*). In addition, several non-invasive reference tissue based approaches have also been used: the simplified reference tissue model (SRTM) and its basis function approach (receptor parametric mapping, RPM), the multilinear reference tissue method (MRTM), and reference Logan (*19–22*). In the present study, the 2T4k_V_b, SRTM and the full reference tissue model (FRTM (*23*)) were examined (Figure 1 provides an overview of the kinetic models used during each step of the analysis). The main aim, however, was to verify

the applicability of a reference tissue model approach given its applicability for large clinical trials.

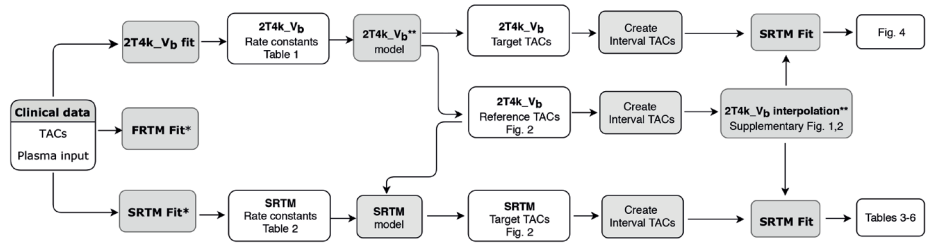


Figure 1. Overview of the kinetic models used during each step of the analysis
* Optimal reference tissue model was selected based on the Akaike information criterion
** Using plasma input

Kinetic parameters for TAC simulations

Both reported whole blood and metabolite corrected arterial plasma input curves were used for [^{18}F]flutemetamol analysis (*24*). For [^{18}F]florbetaben, continuous whole blood curves were generated by scaling the continuous (non-metabolite corrected) plasma curves using discrete whole blood samples. Subsequently, all cortical and cerebellar TACs of all subjects were analysed using the 2T4k_V_b model, SRTM and FRTM (*19,23,25*). The Akaike information criterion was used to determine which reference tissue method best described the kinetics of the tracer (*26*). Both the optimal reference tissue model and the 2T4k_V_b model, were then used for estimating rate constants.

Finally, from the rate constants of the cortical regions (target tissue consisting of anterior and posterior cingulate, frontal, parietal, and lateral and medial temporal cortex) and cerebellum grey matter (reference tissue), mean and standard deviations were calculated and used to establish the range of rate constants for composite cortical and reference tissue regions. The resulting rate constants for the 2T4k_V_b and reference tissue model, can be found in Tables 1 and 2, respectively.

Table 1 Pharmacokinetic parameters for simulating TACs using 2T4k_V_b

	[^{18}F]flutemetamol		[^{18}F]florbetaben	
Level	BP_{ND}	k_3	BP_{ND}	k_3
BP_{ND} I	0.380	0.008	1.000	0.010
BP_{ND} II	0.635	0.013	1.500	0.015
BP_{ND} III	0.890	0.018	2.000	0.020
BP_{ND} IV	1.145	0.023	2.500	0.025
BP_{ND} V	1.400	0.028	3.000	0.030
	BP_{ND}^{NS}	k_3	BP_{ND}^{NS}	k_3
Reference	0.350	0.018	0.950	0.007

BP_{ND} = non-displaceable binding potential of target tissue, V_b had a constant value of 0.05. [^{18}F]flutemetamol target tissue: K_1=0.248, k_2=0.08, reference tissue: K_1=0.32, $k2$=0.103. [^{18}F]florbetaben target tissue: K_1=0.226, k_2=0.069, reference tissue: K_1=0.25, k_2=0.076.

Table 2 Pharmacokinetic parameters for simulating TACs using SRTM

	[^{18}F]flutemetamol	[^{18}F]florbetaben
Level	BP_{ND}	BP_{ND}
BP_{ND} I	0.020	0.021
BP_{ND} II	0.208	0.277
BP_{ND} III	0.397	0.532
BP_{ND} IV	0.585	0.787
BP_{ND} V	0.774	1.042

BP_{ND} = non-displaceable binding potential of target tissue, [^{18}F]flutemetamol: R_1 = 0.775, k_2 = 0.02, [^{18}F]florbetaben: R_1 = 0.904, k_2 = 0.03

TAC simulations

Plasma input generated TACs

Noiseless target and reference tissue TACs of 110 minutes duration were simulated, see Table 1 for kinetic parameters used) using the 2T4k_V_b model to assess the bias in BP_{ND} estimates when fitting these TACs with SRTM. The 2T4k_V_b model was also used to generate a reference tissue TAC for the SRTM simulations described in the next section.

SRTM generated TACs

Using SRTM, tissue target TACs of 110 minutes duration were simulated for the range of BP_{ND} values observed clinically (50 TACs per BP_{ND}, see Table 2 for kinetic parameters

used), along the AD continuum. Various levels of typical PET noise were added to these target TACs only (coefficient of variation (COV) of 1, 2 and 5%, respectively) according to the variance model used by Yaqub et al., creating 50 TACs per noise level for each BP_{ND} (*27*):

$$\sigma_1^2 = \alpha \cdot dcf \cdot dcf \frac{T}{L^2} \quad (1)$$

where σ_1^2 is the variance for each frame, calculated using the whole scanner true counts T, *dcf* the decay correction factor, L the frame length and α a proportionality constant signifying the variance level. In practice, most clinical TACs corresponded best with simulated TACs with 1 or 2% noise added, whilst TACs with 5% noise only corresponded with very small regions with a low BP_{ND} (*7,8*).

Dual-time window protocols

The 'late frame' acquisition window of 90-110 minutes (used within Europe) was left intact, given that it constitutes the approved acquisition protocol for clinical use with visual analysis. Next, data points corresponding to the intervals in the dual-time window protocols were removed from both the target and reference tissue TACs (creating "interval TACs"), ranging from 0 minutes (no interval) to 80 minutes (i.e. interval 10-90 minutes: a 0-10 minutes p.i. acquisition followed by a 90-110 minutes p.i. acquisition) in steps of 10 minutes. This resulted in a total of nine different protocols.

Estimating parameters of interest

Missing data points in the reference tissue TACs, resulting from the introduction of the interval, were interpolated using the 2T4k_V_b model, which was used to fit the interval TACs together with a typical, tracer specific input function, since a well-defined complete reference tissue input curve is required for SRTM to estimate the kinetic parameters of interest (R_1, BP_{ND}, k_2) (*19*). A typical input function could be used for this purpose, based on the observed negligible between-subject variation in the tail of the curve. In future applications of this protocol, either an equivalent approach or an existing population derived input function could be used, provided that a similar injection protocol is used. In addition, boundary values were set for all kinetic parameters (supplementary Tables 1a and b) and for k_3 of the 2T4k_V_b interpolation of the interval (lower boundary: k_3= 0.005).

All TACs were fitted with SRTM, and DVR values were calculated as DVR = BP_{ND}+1. This additional parameter was introduced as it is frequently used to express amyloid

burden in other studies (*28*) and it better allows for expressing any bias in percentages due to its larger values.

Evaluation of outcome parameters

Results of 2T4k_V_b and SRTM generated TACs fitted with SRTM were checked for values that were physiologically not expected, here called outliers (≥ 1.5·max simulated BP_{ND}). These outliers were registered and removed from the overall dataset before further analysis. Subsequently, bias as induced by the interval was assessed for all simulated dual-time window protocols, by calculating the bias between simulated BP_{ND} (BP_{ND}^{sim}) and corresponding mean fitted BP_{ND} (BP_{ND}^{fit}):

$$Bias\ BP_{ND} = BP_{ND}^{fit} - BP_{ND}^{sim} \qquad (2)$$

Next, for each set of 50 simulated TACs at a given BP_{ND}^{sim} and noise level, an acceptability curve was computed by plotting the fraction of data versus a span of bias values ranging from 0 – 1 called the "Absolute bias threshold". These curves were generated for all BP_{ND}^{sim} values across all protocols and the area under the curve (AUC) was calculated for all dual-time window protocols, as a global measure of reliability. Finally, the percentage bias in the interval TAC-derived DVR was assessed by comparing it to DVR derived from the full-kinetic curve:

$$\text{Bias DVR}(\%) = \frac{\text{DVR}_{\text{dual-time window protocol}} - \text{DVR}_{\text{full dataset}}}{\text{DVR}_{\text{full dataset}}} \cdot 100\% \qquad (3)$$

Finally, also the percentage bias in R_1 was assessed by comparing R_1 derived from the various dual-time window protocols with R_1 obtained from the full kinetic curve (in a similar way as for DVR, see equation 3).

Results

Kinetic parameters for TAC simulation

The Akaike information criterion showed that for [^{18}F]flutemetamol (62.7%) and [^{18}F]florbetaben (79.2%) SRTM was the preferred reference tissue method. Pharmacokinetic parameters derived from existing clinical data and used for simulating TACs based on the 2T4k_V_b model and SRTM are presented in Tables 1 and 2. The range of BP_{ND} values is equally spaced and corresponds to the BP_{ND} range present in the

data, with BP_{ND} I being the lowest and BP_{ND} V the highest value. Of note, as described previously, BP_{ND} estimates are always different between 2T4k_V_b and SRTM. In the present study, additional differences are present, since the first corresponds to the sum of specific binding and a slow component of non-specific binding (in the target tissue), while in the latter a correction for all non-specific binding is made, provided it is the same in target and reference tissues (*18,28,29*).

TAC Simulation

Full reference and target tissue TACs, the latter covering the range of BP_{ND}^{sim} values, are shown in Figure 2. Figure 3 shows the pattern in which the different noise levels were simulated for a target tissue TAC (global cortical region), resembling the shape of clinical TACs published previously (*7,8*).

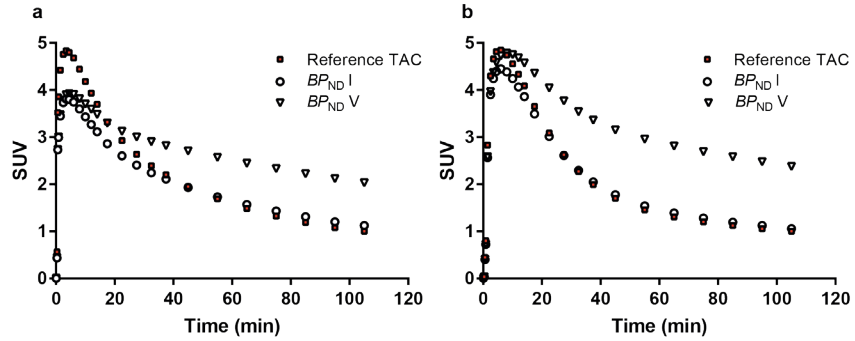


Figure 2. SRTM simulated target tissue TACs for the range of BP_{ND}^{sim} values and 2T4k_V_b generated reference tissue TACs for (a) [^{18}F]flutemetamol and **(b)** [^{18}F]florbetaben

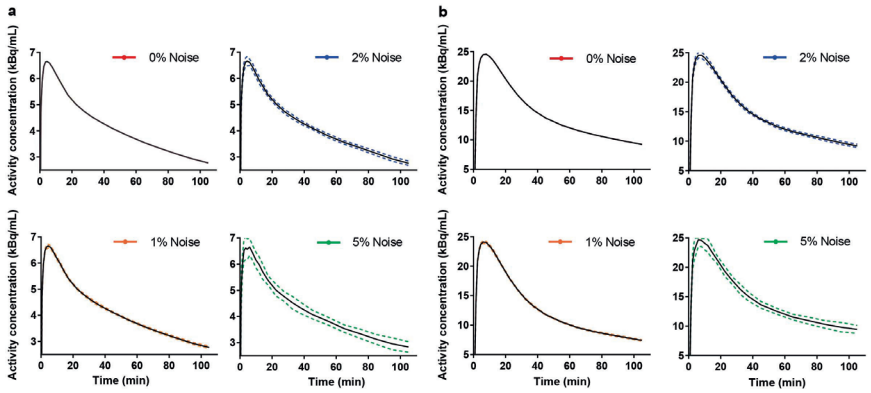


Figure 3. SRTM simulated TACs from a global cortical region (mean TAC value ± SD, shown solid and dashed lines, respectively) for all noise levels (COV 0–5%) for **(a)** [^{18}F]flutemetamol and **(b)** [^{18}F]florbetaben. All TACs were simulated using a BP_{ND}^{sim} III (Table 2).

Estimating parameters of interest

Plasma input generated TACs

As can be seen in Figure 4, SRTM-derived DVR resulted in a systematic bias when fitting the noiseless, full-kinetic curve (90-90) 2T4k_V_b generated TACs for both tracers. For [^{18}F]flutemetamol this bias ranged between the 0.17 and 1.95% and for [^{18}F]florbetaben between the 2.62 and 6.04%. Compared to the full-kinetic curves, 2T4k_V_b interval TACs showed a greater bias in SRTM-derived DVR only for the 10-90 and 20-90 interval TACs (maximum bias of 3.10 and 2.25%, respectively for [^{18}F]flutemetamol and maximum bias of 8.73 and 10.10%, respectively for [^{18}F]florbetaben) and comparable or smaller bias for the other interval TACs.

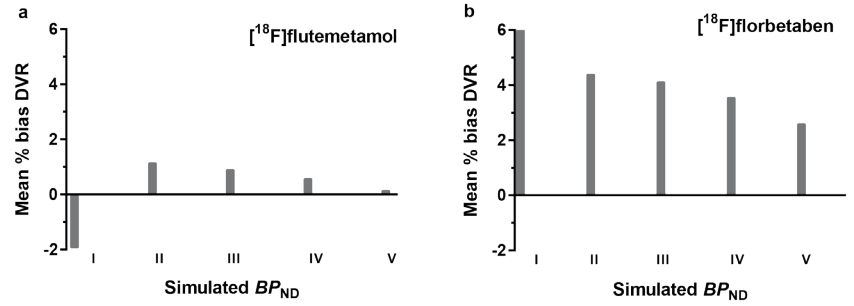


Figure 4. Percentage bias in SRTM-derived DVR of 2T4k_V_b generated TACs. (a) Bias in DVR for [^{18}F]flutemetamol. **(b)** Bias in DVR for [^{18}F]florbetaben.

Evaluation of outcome parameter

Outliers

No outliers were present when no noise was added to the TACs and the largest number of outliers occurred at the highest noise level for both tracers (Tables 3 and 4). More specifically, a positive correlation was observed between percentage of outliers and interval (R^2=0.47, p=0.04*)* and between percentage outliers and noise levels (R^2=0.91, p<0.05*)* for [^{18}F]flutemetamol. For [^{18}F]florbetaben, outliers were mostly present at a noise level of 5% and the number increased for higher BP_{ND}^{sim} values (R^2=0.85, p=0.03*)*. The percentage of [^{18}F]flutemetamol outliers exceeded the 20% for the two largest intervals, while for [^{18}F]florbetaben it exceeded the 20% only for the 10-90 interval.

Table 3. Bias in DVR and outliers as a result of noise and the dual-time window protocol for SRTM generated [^{18}F]flutemetamol TACs

FLUT		% Bias DVR			% Outlier		
	Interval	1%	2%	5%	1%	2%	5%
BP_{ND} I	10-90	0.0	1.9	2.0	0.0	0.0	0.0
	20-90	6.4	6.1	2.4	0.0	0.0	0.0
	30-90	0.5	0.5	3.1	0.0	0.0	0.0
	40-90	1.5	1.6	1.6	0.0	0.0	0.0
	50-90	1.5	1.5	1.7	0.0	0.0	0.0
	60-90	1.5	1.6	1.9	0.0	0.0	0.0
	70-90	1.8	2.1	0.3	0.0	0.0	0.0
	80-90	0.8	1.3	1.3	0.0	0.0	0.0
	90-90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} II	10-90	1.0	6.2	8.2	0.0	2.0	12.0
	20-90	1.5	0.2	4.0	0.0	0.0	2.0
	30-90	0.1	0.2	1.3	0.0	0.0	0.0
	40-90	0.0	0.3	0.1	0.0	0.0	0.0
	50-90	0.0	0.3	0.4	0.0	0.0	0.0
	60-90	0.0	0.2	0.6	0.0	0.0	0.0
	70-90	0.0	0.1	0.6	0.0	0.0	0.0
	80-90	0.1	0.0	0.3	0.0	0.0	0.0
	90-90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} III	10-90	0.9	1.4	2.0	0.0	10.0	34.0
	20-90	0.1	0.5	0.6	0.0	0.0	10.0
	30-90	0.1	0.3	0.3	0.0	0.0	4.0
	40-90	0.1	0.5	0.0	0.0	0.0	4.0
	50-90	0.1	0.5	0.2	0.0	0.0	4.0
	60-90	0.1	0.5	0.1	0.0	0.0	4.0
	70-90	0.1	0.4	0.3	0.0	0.0	4.0
	80-90	0.0	0.1	0.5	0.0	0.0	2.0
	90-90	0.0	0.0	0.0	0.0	0.0	2.0
BP_{ND} IV	10-90	1.6	2.7	1.8	0.0	2.0	32.0
	20-90	0.2	0.3	0.8	0.0	0.0	12.0
	30-90	0.0	0.0	1.3	0.0	0.0	6.0
	40-90	0.0	0.1	0.6	0.0	0.0	0.0
	50-90	0.0	0.1	0.7	0.0	0.0	0.0
	60-90	0.0	0.1	0.7	0.0	0.0	0.0
	70-90	0.1	0.2	0.7	0.0	0.0	0.0
	80-90	0.0	0.1	0.9	0.0	0.0	0.0
	90-90	0.0	0.0	0.0	0.0	0.0	0.0

Table 3. Bias in DVR and outliers as a result of noise and the dual-time window protocol for SRTM generated [^{18}F]flutemetamol TACs (continued)

FLUT		% Bias DVR			% Outlier		
	Interval	1%	2%	5%	1%	2%	5%
BP_{ND} V	10-90	1.2	0.1	6.5	2.0	14.0	48.0
	20-90	0.1	0.0	0.5	0.0	0.0	22.0
	30-90	0.0	0.6	0.8	0.0	0.0	18.0
	40-90	0.0	0.5	1.0	0.0	0.0	16.0
	50-90	0.0	0.6	0.9	0.0	0.0	14.0
	60-90	0.0	0.6	0.3	0.0	0.0	12.0
	70-90	0.0	0.4	0.5	0.0	0.0	14.0
	80-90	0.0	0.2	0.4	0.0	0.0	12.0
	90-90	0.0	0.0	0.0	0.0	0.0	12.0

Bias in simulated DVR compared to full-kinetic curve DVR and % outliers across noise levels (1-5%) for [^{18}F]flutemetamol.

Table 4. Bias in DVR and outliers as a result of noise and the dual-time window protocol for SRTM generated [^{18}F]florbetaben TACs

FLUT		% Bias DVR			% Outlier		
	Interval	1%	2%	5%	1%	2%	5%
BP_{ND} I	10 90	2.1	0.9	3.3	0.0	0.0	0.0
	20 90	0.9	0.2	2.3	0.0	0.0	0.0
	30 90	1.0	1.2	1.8	0.0	0.0	0.0
	40 90	0.6	1.9	0.3	0.0	0.0	0.0
	50 90	0.1	1.1	1.2	0.0	0.0	0.0
	60 90	1.0	1.0	0.4	0.0	0.0	0.0
	70 90	1.0	1.1	0.7	0.0	0.0	0.0
	80 90	0.5	0.4	0.4	0.0	0.0	0.0
	90 90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} II	10 90	0.1	3.4	9.1	0.0	0.0	0.0
	20 90	0.1	0.1	5.0	0.0	0.0	0.0
	30 90	0.1	0.4	0.2	0.0	0.0	0.0
	40 90	0.0	0.2	0.1	0.0	0.0	0.0
	50 90	0.0	0.2	0.3	0.0	0.0	0.0
	60 90	0.1	0.1	0.3	0.0	0.0	0.0
	70 90	0.0	0.1	0.3	0.0	0.0	0.0
	80 90	0.0	0.0	0.2	0.0	0.0	0.0
	90 90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} III	10 90	0.2	0.7	3.5	0.0	2.0	16.0
	20 90	0.2	0.2	1.8	0.0	0.0	0.0

Table 4. Bias in DVR and outliers as a result of noise and the dual-time window protocol for SRTM generated [^{18}F]florbetaben TACs (continued)

FLUT		% Bias DVR			% Outlier		
	Interval	1%	2%	5%	1%	2%	5%
	30 90	0.3	0.5	0.6	0.0	0.0	0.0
	40 90	0.2	0.4	0.7	0.0	0.0	0.0
	50 90	0.1	0.4	0.7	0.0	0.0	0.0
	60 90	0.1	0.4	0.8	0.0	0.0	0.0
	70 90	0.0	0.3	0.5	0.0	0.0	0.0
	80 90	0.0	0.1	0.3	0.0	0.0	0.0
	90 90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} IV	10 90	0.5	0.8	4.0	0.0	0.0	20.0
	20 90	0.0	0.2	1.3	0.0	0.0	0.0
	30 90	0.2	0.3	0.1	0.0	0.0	0.0
	40 90	0.0	0.2	0.2	0.0	0.0	0.0
	50 90	0.0	0.1	0.3	0.0	0.0	0.0
	60 90	0.0	0.1	0.3	0.0	0.0	0.0
	70 90	0.1	0.1	0.3	0.0	0.0	0.0
	80 90	0.0	0.0	0.5	0.0	0.0	0.0
	90 90	0.0	0.0	0.0	0.0	0.0	0.0
BP_{ND} V	10 90	0.4	1.4	2.7	0.0	0.0	32.0
	20 90	0.3	0.5	0.9	0.0	0.0	10.0
	30 90	0.3	0.8	0.7	0.0	0.0	2.0
	40 90	0.0	0.6	0.4	0.0	0.0	0.0
	50 90	0.0	0.6	0.8	0.0	0.0	2.0
	60 90	0.0	0.6	0.6	0.0	0.0	2.0
	70 90	0.0	0.4	0.2	0.0	0.0	2.0
	80 90	0.0	0.2	0.0	0.0	0.0	2.0
	90 90	0.0	0.0	0.0	0.0	0.0	0.0

Bias in simulated DVR compared to full-kinetic curve DVR and % outliers across noise levels (1-5%) for [^{18}F]florbetaben.

SRTM generated [^{18}F]flutemetamol interval TACs

The interpolated TACs can be found in supplementary Figure 1. Table 5 shows, as expected, increasing absolute bias in BP_{ND}^{fit}, with increasing noise levels (R^2= 0.98, p=0.01) and longer intervals (R^2= 0.51, p=0.03*)* (maximum bias of 0.128 for the 10-90 BP_{ND}^{sim} II interval, at 5% noise). In addition, there was a trend towards a negative correlation between absolute bias and BP_{ND}^{sim} (R^2=0.74, p=0.06*)*. The AUC values calculated from the acceptability curves show a trend of smaller (poorer) AUC values at higher noise levels (R^2=1.00, p<0.001) and larger intervals (R^2=0.64, p=0.01). As expected, the full-kinetic curve provided the highest AUC values, except for the lowest BP_{ND}^{sim}. Furthermore, the

10-90 and 20-90 minutes intervals result in a bias in DVR of maximal 6.4%, while all other intervals show a bias in DVR of maximum 1.6% for noise levels of up to 2%. For higher noise levels corresponding to very small regions (5%), bias in DVR was a maximum of 8.2 and 4.0%, for the 10-90 and 20-90 intervals (Table 3). Finally, Figure 5 shows the percentage bias in R_1, which was only larger than 1% for the largest interval. The bias in R_1 estimates also increased with increasing noise level (R^2=0.99, p=0.004, COV2: ranging from 0.04 to -0.497 and COV5: 0.034 to -3.462).

Table 5. Absolute bias in BP_{ND} and AUC as a result of noise and the dual-time window protocol for SRTM generated [^{18}F]flutemetamol TACs

FLUT		Absolute mean bias BP_{ND} (SD)				AUC			
	Interval	0%	1%	2%	5%	0%	1%	2%	5%
BP_{ND} I	10-90	0.001 (0.000)	0.017 (0.058)	0.042 (0.079)	0.093 (0.154)	0.993	0.971	0.943	0.894
	20-90	0.001 (0.000)	0.084 (0.283)	0.086 (0.276)	0.098 (0.211)	0.993	0.908	0.902	0.887
	30-90	0.001 (0.000)	0.012 (0.079)	0.017 (0.074)	0.105 (0.222)	0.993	0.969	0.961	0.878
	40-90	0.001 (0.000)	0.002 (0.011)	0.006 (0.022)	0.055 (0.113)	0.993	0.986	0.977	0.929
	50-90	0.000 (0.000)	0.002 (0.011)	0.007 (0.025)	0.053 (0.116)	0.993	0.986	0.976	0.929
	60-90	0.000 (0.000)	0.002 (0.029)	0.006 (0.045)	0.051 (0.115)	0.993	0.981	0.974	0.932
	70-90	0.000 (0.000)	0.036 (0.119)	0.000 (0.020)	0.075 (0.147)	0.993	0.948	0.978	0.909
	80-90	0.000 (0.000)	0.026 (0.174)	0.035 (0.181)	0.058 (0.133)	1.000	0.955	0.946	0.923
	90-90	0.000 (0.000)	0.018 (0.162)	0.022 (0.162)	0.072 (0.202)	1.000	0.962	0.958	0.91
BP_{ND} II	10-90	0.002 (0.000)	0.014 (0.046)	0.081 (0.217)	0.128 (0.271)	0.993	0.97	0.895	0.838
	20-90	0.001 (0.000)	0.021 (0.135)	0.009 (0.040)	0.076 (0.200)	0.993	0.963	0.965	0.881
	30-90	0.001 (0.000)	0.002 (0.015)	0.003 (0.030)	0.043 (0.159)	0.993	0.982	0.971	0.906
	40-90	0.000 (0.000)	0.002 (0.015)	0.003 (0.029)	0.029 (0.112)	0.993	0.982	0.971	0.919
	50-90	0.000 (0.000)	0.002 (0.015)	0.002 (0.029)	0.022 (0.097)	0.993	0.982	0.972	0.923
	60-90	0.000 (0.000)	0.003 (0.014)	0.004 (0.029)	0.020 (0.095)	0.993	0.983	0.972	0.922
	70-90	0.000 (0.000)	0.002 (0.014)	0.005 (0.029)	0.020 (0.096)	0.993	0.984	0.972	0.924
	80-90	0.000 (0.000)	0.002 (0.013)	0.007 (0.027)	0.023 (0.098)	1.000	0.985	0.973	0.924
	90-90	0.000 (0.000)	0.003 (0.013)	0.006 (0.024)	0.027 (0.098)	1.000	0.985	0.975	0.925
BP_{ND} III	10-90	0.003 (0.000)	0.014 (0.063)	0.031 (0.099)	0.023 (0.128)	0.993	0.956	0.924	0.898
	20-90	0.002 (0.000)	0.001 (0.026)	0.019 (0.080)	0.06 (0.145)	0.993	0.975	0.946	0.876
	30-90	0.001 (0.000)	0.001 (0.023)	0.007 (0.054)	0.056 (0.158)	0.993	0.977	0.956	0.881
	40-90	0.001 (0.000)	0.001 (0.023)	0.004 (0.049)	0.051 (0.146)	0.993	0.978	0.958	0.885
	50-90	0.001 (0.000)	0.001 (0.023)	0.005 (0.048)	0.049 (0.147)	0.993	0.978	0.959	0.886
	60-90	0.001 (0.000)	0.001 (0.022)	0.005 (0.048)	0.05 (0.147)	0.993	0.978	0.959	0.887
	70-90	0.001 (0.000)	0.000 (0.021)	0.006 (0.047)	0.047 (0.144)	0.993	0.979	0.96	0.888
	80-90	0.000 (0.000)	0.001 (0.022)	0.01 (0.046)	0.059 (0.164)	1.000	0.979	0.961	0.883
	90-90	0.000 (0.000)	0.001 (0.021)	0.012 (0.043)	0.051 (0.142)	1.000	0.979	0.963	0.894
BP_{ND} IV	10-90	0.005 (0.000)	0.034 (0.082)	0.041 (0.147)	0.027 (0.192)	0.993	0.932	0.892	0.854
	20-90	0.003 (0.000)	0.012 (0.038)	0.004 (0.067)	0.043 (0.183)	0.993	0.963	0.942	0.859
	30-90	0.002 (0.000)	0.008 (0.032)	0.002 (0.055)	0.036 (0.168)	0.993	0.968	0.949	0.866
	40-90	0.001 (0.000)	0.008 (0.030)	0.003 (0.054)	0.048 (0.173)	0.993	0.97	0.95	0.858
	50-90	0.001 (0.000)	0.007 (0.031)	0.003 (0.054)	0.045 (0.174)	0.993	0.97	0.949	0.858
	60-90	0.001 (0.000)	0.008 (0.030)	0.003 (0.053)	0.046 (0.169)	0.993	0.969	0.951	0.863
	70-90	0.001 (0.000)	0.007 (0.031)	0.004 (0.051)	0.045 (0.165)	0.993	0.969	0.953	0.868
	80-90	0.000 (0.000)	0.008 (0.030)	0.000 (0.050)	0.043 (0.163)	1.000	0.969	0.954	0.867
	90-90	0.000 (0.000)	0.008 (0.028)	0.001 (0.045)	0.057 (0.179)	1.000	0.97	0.959	0.861
BP_{ND} V	10-90	0.006 (0.000)	0.023 (0.082)	0.005 (0.145)	0.117 (0.111)	0.988	0.929	0.885	0.852
	20-90	0.004 (0.000)	0.001 (0.042)	0.004 (0.090)	0.011 (0.179)	0.993	0.96	0.925	0.85
	30-90	0.002 (0.000)	0.000 (0.037)	0.014 (0.074)	0.015 (0.179)	0.993	0.965	0.938	0.851
	40-90	0.001 (0.000)	0.001 (0.034)	0.014 (0.072)	0.019 (0.168)	0.993	0.968	0.94	0.853
	50-90	0.001 (0.000)	0.001 (0.034)	0.015 (0.071)	0.018 (0.164)	0.993	0.968	0.94	0.855
	60-90	0.001 (0.000)	0.001 (0.034)	0.015 (0.071)	0.008 (0.173)	0.993	0.968	0.94	0.849
	70-90	0.001 (0.000)	0.001 (0.033)	0.012 (0.072)	0.01 (0.163)	0.993	0.968	0.939	0.856
	80-90	0.000 (0.000)	0.001 (0.031)	0.008 (0.067)	0.006 (0.172)	1.000	0.969	0.943	0.849
	90-90	0.000 (0.000)	0.001 (0.029)	0.004 (0.061)	0.002 (0.150)	1.000	0.971	0.949	0.867

Absolute bias in BP_{ND}^{sim} and the area under the curve (AUC) for the acceptability curves across all noise levels (0-5%) for [^{18}F]flutemetamol.

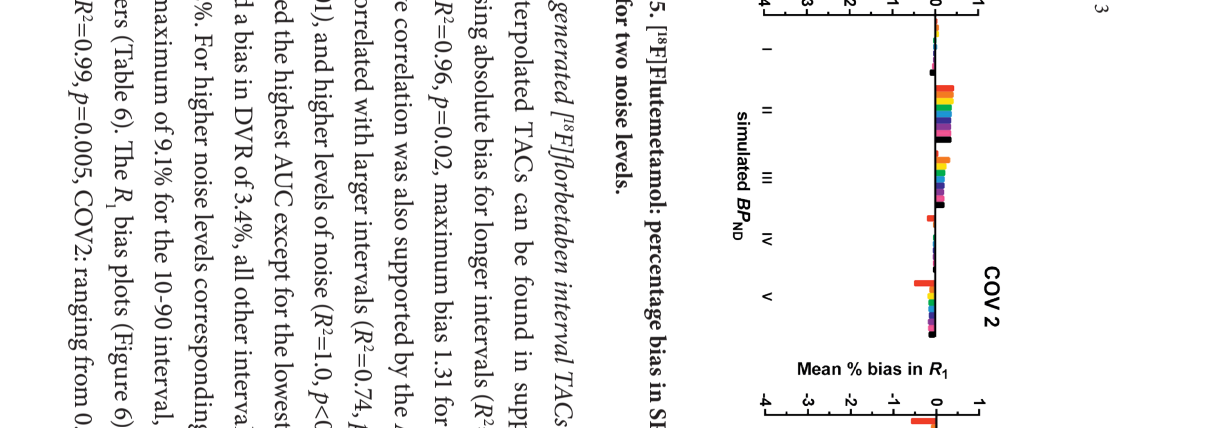


Figure 5. [^{18}F]Flutemetamol: percentage bias in SRTM-derived R_1 across the range of BP_{ND}^{sim} values for two noise levels.

SRTM generated [^{18}F]florbetaben interval TACs

The interpolated TACs can be found in supplementary Figure 2. Table 6 shows an increasing absolute bias for longer intervals (R^2=0.48, p=0.04) as well as for higher noise levels (R^2=0.96, p=0.02, maximum bias 1.31 for the 10-90 BP_{ND}^{sim} II interval, at 5%). This positive correlation was also supported by the AUC values, where lower (poorer) values were correlated with larger intervals (R^2=0.74, p=0.002), higher BP_{ND}^{sim} values (R^2=0.98, p=0.001), and higher levels of noise (R^2=1.0, p<0.001). As expected, the full-kinetic curve provided the highest AUC except for the lowest BP_{ND}^{sim}. Furthermore, the 10-90 interval showed a bias in DVR of 3.4%, all other intervals showed a bias of ≤ 1.9% for noise levels up to 2%. For higher noise levels corresponding to very small regions (5%), bias in DVR was a maximum of 9.1% for the 10-90 interval, 5.0% for the 20-90 interval and 1.8% for all others (Table 6). The R_1 bias plots (Figure 6) show an increase in bias at higher noise levels (R^2=0.99, p=0.005, COV2: ranging from 0.001 to -0.359 and COV5: 0.003 to -1.497).

Table 6. Absolute bias in BP_{ND} and AUC as a result of noise and the dual-time-window protocol for SRTM generated [^{18}F]florbetaben TACs

FBB		Absolute mean bias BP_{ND} (SD)				AUC			
	Interval	0%	1%	2%	5%	0%	1%	2%	5%
BP_{ND} I	10-90	0.000 (0.000)	0.012 (0.026)	0.021 (0.043)	0.059 (0.080)	0.993	0.979	0.966	0.929
	20-90	0.000 (0.000)	0.000 (0.011)	0.013 (0.040)	0.048 (0.076)	0.993	0.986	0.971	0.937
	30-90	0.000 (0.000)	0.000 (0.008)	0.001 (0.028)	0.006 (0.046)	0.993	0.988	0.979	0.968
	40-90	0.000 (0.000)	0.003 (0.032)	0.008 (0.014)	0.021 (0.063)	0.993	0.978	0.981	0.956
	50-90	0.000 (0.000)	0.009 (0.012)	0.000 (0.037)	0.012 (0.056)	0.993	0.983	0.978	0.962
	60-90	0.000 (0.000)	0.000 (0.007)	0.001 (0.015)	0.029 (0.063)	0.993	0.989	0.983	0.956
	70-90	0.000 (0.000)	0.001 (0.006)	0.000 (0.013)	0.032 (0.069)	0.993	0.99	0.984	0.953
	80-90	0.000 (0.000)	0.005 (0.036)	0.007 (0.066)	0.029 (0.081)	1.000	0.978	0.963	0.948
	90-90	0.000 (0.000)	0.01 (0.011)	0.011 (0.064)	0.024 (0.067)	1.000	0.983	0.962	0.956
BP_{ND} II	10-90	0.001 (0.000)	0.003 (0.013)	0.045 (0.167)	0.131 (0.307)	0.993	0.984	0.935	0.861
	20-90	0.001 (0.000)	0.000 (0.012)	0.000 (0.023)	0.079 (0.207)	0.993	0.985	0.976	0.891
	30-90	0.002 (0.000)	0.000 (0.012)	0.003 (0.023)	0.016 (0.063)	0.993	0.985	0.975	0.946
	40-90	0.000 (0.000)	0.001 (0.012)	0.001 (0.023)	0.012 (0.062)	0.993	0.985	0.976	0.946
	50-90	0.000 (0.000)	0.001 (0.011)	0.001 (0.022)	0.01 (0.060)	0.993	0.986	0.977	0.948
	60-90	0.000 (0.000)	0.002 (0.010)	0.000 (0.021)	0.009 (0.056)	0.993	0.987	0.978	0.948
	70-90	0.000 (0.000)	0.001 (0.009)	0.001 (0.020)	0.01 (0.055)	0.993	0.988	0.979	0.95
	80-90	0.000 (0.000)	0.001 (0.007)	0.002 (0.017)	0.011 (0.053)	1.000	0.989	0.981	0.954
	90-90	0.000 (0.000)	0.001 (0.008)	0.002 (0.015)	0.014 (0.053)	1.000	0.989	0.983	0.954
BP_{ND} III	10-90	0.002 (0.000)	0.002 (0.022)	0.014 (0.066)	0.084 (0.159)	0.993	0.978	0.955	0.875
	20-90	0.002 (0.000)	0.003 (0.016)	0.000 (0.041)	0.058 (0.117)	0.993	0.982	0.964	0.898
	30-90	0.003 (0.000)	0.004 (0.016)	0.004 (0.035)	0.038 (0.112)	0.993	0.982	0.966	0.91
	40-90	0.001 (0.000)	0.002 (0.016)	0.003 (0.036)	0.041 (0.116)	0.993	0.982	0.966	0.909

Table 6. Absolute bias in BP_{ND} and AUC as a result of noise and the dual-time-window protocol for SRTM generated [^{18}F]florbetaben TACs (continued)

FBB		Absolute mean bias BP_{ND} (SD)				AUC			
	Interval	0%	1%	2%	5%	0%	1%	2%	5%
	50-90	0.001 (0.000)	0.002 (0.016)	0.002 (0.035)	0.041 (0.117)	0.993	0.982	0.967	0.908
	60-90	0.001 (0.000)	0.001 (0.015)	0.002 (0.034)	0.041 (0.114)	0.993	0.983	0.968	0.909
	70-90	0.001 (0.000)	0.001 (0.014)	0.001 (0.031)	0.037 (0.108)	0.993	0.984	0.971	0.911
	80-90	0.000 (0.000)	0.000 (0.014)	0.002 (0.030)	0.033 (0.092)	1.000	0.985	0.972	0.926
	90-90	0.000 (0.000)	0.000 (0.013)	0.003 (0.027)	0.029 (0.083)	1.000	0.985	0.974	0.933
BP_{ND} IV	10-90	0.003 (0.000)	0.014 (0.035)	0.014 (0.074)	0.085 (0.184)	0.993	0.964	0.943	0.848
	20-90	0.004 (0.000)	0.004 (0.024)	0.003 (0.047)	0.036 (0.148)	0.993	0.975	0.956	0.891
	30-90	0.005 (0.000)	0.001 (0.023)	0.006 (0.046)	0.011 (0.115)	0.993	0.976	0.958	0.904
	40-90	0.001 (0.000)	0.004 (0.023)	0.003 (0.046)	0.008 (0.108)	0.993	0.976	0.958	0.907
	50-90	0.001 (0.000)	0.004 (0.024)	0.002 (0.045)	0.006 (0.107)	0.993	0.975	0.959	0.909
	60-90	0.001 (0.000)	0.005 (0.022)	0.002 (0.043)	0.007 (0.099)	0.993	0.976	0.961	0.917
	70-90	0.001 (0.000)	0.004 (0.022)	0.003 (0.040)	0.006 (0.092)	0.993	0.977	0.964	0.924
	80-90	0.000 (0.000)	0.005 (0.021)	0.000 (0.037)	0.003 (0.089)	1.000	0.977	0.966	0.926
	90-90	0.000 (0.000)	0.005 (0.019)	0.000 (0.034)	0.012 (0.091)	1.000	0.978	0.97	0.925
BP_{ND} V	10-90	0.005 (0.000)	0.005 (0.044)	0.024 (0.122)	0.029 (0.111)	0.993	0.96	0.911	0.903
	20-90	0.006 (0.000)	0.008 (0.027)	0.015 (0.056)	0.008 (0.147)	0.988	0.971	0.95	0.882
	30-90	0.007 (0.000)	0.008 (0.027)	0.021 (0.051)	0.011 (0.165)	0.988	0.972	0.954	0.87
	40-90	0.002 (0.000)	0.002 (0.026)	0.016 (0.052)	0.018 (0.172)	0.993	0.973	0.955	0.868
	50-90	0.002 (0.000)	0.002 (0.026)	0.016 (0.052)	0.010 (0.156)	0.993	0.973	0.955	0.878
	60-90	0.001 (0.000)	0.002 (0.025)	0.016 (0.051)	0.015 (0.156)	0.993	0.974	0.955	0.875
	70-90	0.001 (0.000)	0.002 (0.024)	0.012 (0.050)	0.022 (0.153)	0.993	0.976	0.955	0.876
	80-90	0.000 (0.000)	0.002 (0.022)	0.008 (0.045)	0.027 (0.147)	0.993	0.978	0.959	0.876
	90-90	0.000 (0.000)	0.002 (0.020)	0.005 (0.041)	0.026 (0.135)	0.993	0.979	0.963	0.888

Absolute bias in $BP_{\mathrm{ND}}^{\mathrm{sim}}$ and the area under the curve (AUC) for the acceptability curves across all noise levels (0-5%) for [^{18}F]florbetaben.

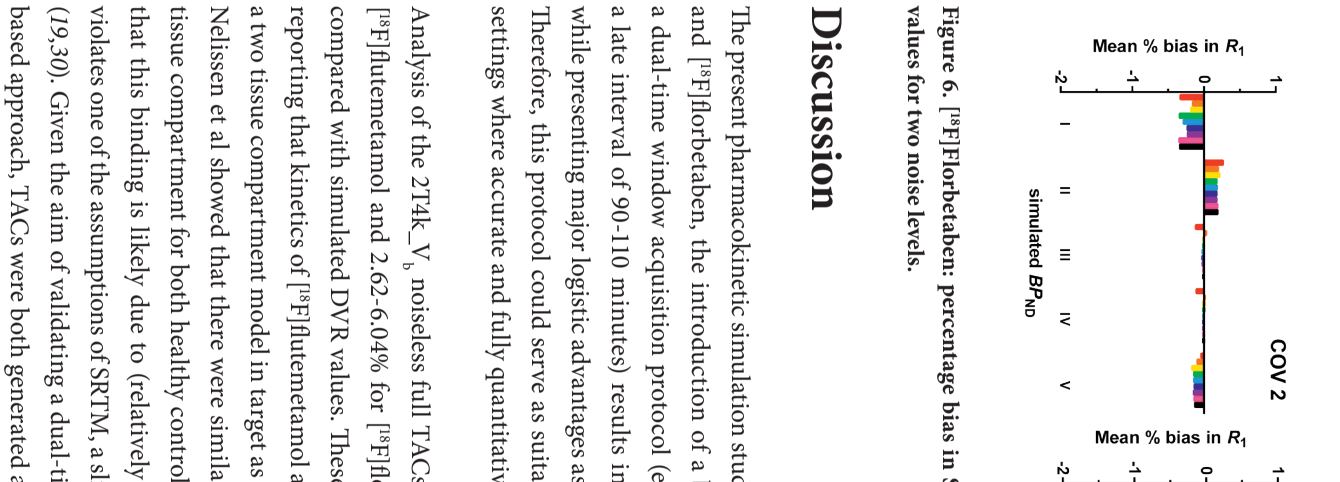


Figure 6. [^{18}F]Florbetaben: percentage bias in SRTM-derived R_1 across the range of $BP_{\mathrm{ND}}^{\mathrm{sim}}$ values for two noise levels.

Discussion

The present pharmacokinetic simulation study demonstrated that, for [^{18}F]flutemetamol and [^{18}F]florbetaben, the introduction of a break with a maximum of 60 minutes in a dual-time window acquisition protocol (early interval of 0-30 minutes followed by a late interval of 90-110 minutes) results in a minimal loss in quantitative accuracy while presenting major logistic advantages as compared with full dynamic acquisitions. Therefore, this protocol could serve as suitable alternative in research or clinical trial settings where accurate and fully quantitative measurements might be required.

Analysis of the 2T4k_V_b noiseless full TACs, showed a systematic bias (0.17-1.95% for [^{18}F]flutemetamol and 2.62-6.04% for [^{18}F]florbetaben) in SRTM-derived DVR values compared with simulated DVR values. These findings are in line with previous studies reporting that kinetics of [^{18}F]flutemetamol and [^{18}F]florbetaben are better described by a two tissue compartment model in target as well as reference tissues (*8*,*18*). In addition, Nelissen et al showed that there were similar levels of binding in the second reference tissue compartment for both healthy control and AD subjects and therefore concluded that this binding is likely due to (relatively slow) non-specific retention (*18*). As this violates one of the assumptions of SRTM, a slight bias in DVR estimates can be expected (*19*,*30*). Given the aim of validating a dual-time window protocol for a reference tissue based approach, TACs were both generated and fitted according to the SRTM model to prevent a systematic bias in the results.

A first examination of SRTM-derived BP_{ND}^{fit} values revealed that most outliers were observed for fits of the 10-90 minutes interval, and, to a lesser extent, also for fits of the 20-90 minutes interval. In addition, compared with the full-kinetic curve, the bias in DVR only exceeded previously reported [^{11}C]PiB TRT values, of which [^{18}F]flutemetamol is an analogue, for the 10-90 and 20-90 minutes intervals for [^{18}F]flutemetamol (*9*,*10*). For the other dual-time window protocols the bias remained ≤ 3.1%. Analysis of [^{18}F]florbetaben data showed a bias of ≤ 9.1% for the 10-90 minutes interval and ≤ 5.0% for the 20-90 interval for the highest noise levels, while for all intervals it was ≤ 1.9%. The latter well within previously reported TRT values for [^{18}F]florbetaben SUVR data (ranging between 2.9% HC and 6.2% AD) (*31*). Reported AUC values also showed a general trend of worse values for longer breaks and higher noise level, with the exception of some extremely low BP_{ND} cases, where the performance of SRTM is known to be suboptimal (*10*).

Finally, the bias plots of SRTM generated TACs demonstrated that bias in SRTM-derived R_1 increased as a function of noise and interval for both tracers. More specifically, a larger error in R_1 (>3% for [^{18}F]flutemetamol and >1% for [^{18}F]florbetaben) was observed for the 10-90 minutes intervals compared with the other intervals. For practical applications this error would be negligible since the TRT of flow is known to be approximately 9% (*32*). As expected, the results showed that the length of the interval is related to bias in BP_{ND}^{fit} or DVR and the number of outliers. More specifically, results suggest that it is not advisable to use the 10-90 and 20-90 minutes intervals for full quantification, especially due to the relatively large percentage of outliers and larger bias in DVR compared with other intervals. Moreover, the observed larger amount of unusable data would result in smaller power to detect changes in clinical trials.

Shorter scan durations are better for the patient and, as such, longer breaks would be preferred. Since the 10-90 and 20-90 minutes intervals result in a large number of outliers and larger bias, the 30-90 minutes interval would be a good compromise. This interval would have the additional advantage of a 60 minutes break, which may allow for interleaved scanning protocols. Consequently, the 30-90 minutes interval is recommended as the optimal trade-off between patient comfort and quantitative accuracy (bias in DVR < 2% and a maximum of 18 % outliers for highest noise level and BP_{ND}^{fit}). This conclusion is in agreement with recent work of Bullich et al regarding the optimal [^{18}F]florbetaben dual-time window protocol (*17*). Based on their analysis of clinical data, which did not include the 90-110 minutes diagnostic window, a dual-time window protocol of 0-30 and 120-140 minutes was described as optimal. However, their

simulations also supported that 0-30 and 90-110 minutes scanning times would maintain the best compromise between quantitative accuracy and patient comfort. The present simulation study, including TACs representing the AD spectrum and different noise levels, further validated their findings.

A major advantage of a 60 minutes gap in the scanning protocol is that it allows for interleaved scanning protocols, in which the first scan of the second patient can be acquired within the resting period (interval) of the first patient. An interleaved scanning protocol would increase both patient throughput and efficient use of tracer batches, thereby decreasing costs. An assessment of the practical feasibility of such an interleaved scanning protocol is beyond the scope of the present study and needs to be addressed in future studies.

Main limitations of the current study include the use of fixed K_1 and k_2 parameters for simulations, the limited sample size of the available clinical dataset, and the extrapolation of TRT variability from other radiotracers to this work. The first limits the possibility of assessing the impact of changes in cerebral blood flow on dual-time window protocol based quantification, but it can be expected to introduce only small additional bias over and above the one introduced by the protocol itself (*11*). Regarding the second, additional clinical data would have allowed the verification of the simulation results, which remains a goal for future work once larger cohorts are available. With respect to extrapolating TRT variability, although values from other tracers might not directly translate to our data, they are expected to be in comparable ranges (*24*). In addition, although outside of the scope of this study, the evaluation of parametric methods for quantification of dual-time window-derived data is warranted, which would require imaging data in order to optimize image contrast and reduce noise and artifacts. Finally, it must be noted that the goal of this study was not to compare these two tracers, but to identify the optimal dynamic dual-time window scanning protocol for both of them. In order to make a head-to-head comparison between tracers, PET imaging data from both tracers within the same patient would be required.

Conclusion

Accurate estimates of BP_{ND}^{fit} can be obtained for both [^{18}F]flutemetamol and [^{18}F]florbetaben using a 60 minutes dual-time window protocol, with dynamic scanning from 0 to 30 and again from 90 to 110 minutes. This protocol results in a limited number of outliers, and an acceptable bias in BP_{ND}^{fit} and DVR estimates. Moreover, it enables interleaved scanning protocols, optimizing tracer batch usage and patient throughput, thereby reducing costs and improving patient comfort.

Acknowledgements

We would like to thank the staff of the department of Nuclear Medicine, University of Leipzig for skilful acquisition and analyses of the [^{18}F]florbetaben data during the clinical development.

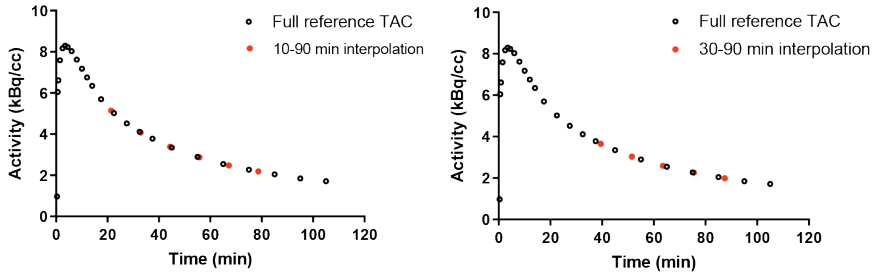
References

1. Braak H, Braak E. 9 Springer-Verlag 1991 Review Neuropathological stageing of Alzheimer-related changes. 1991.
2. Thal DR, Rüb U, Orantes M, Braak H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology. 2002;58:1791-1800.
3. Alzheimer's Association. 2014 Alzheimer's disease facts and figures. Alzheimers Dement. 2014;10:e47-e92.
4. Jack CR, Knopman DS, Jagust WJ, et al. Update on hypothetical model of Alzheimer's disease biomarkers. Lancet Neurol. 2013;12:207-216.
5. Mallik A, Drzezga A, Minoshima S. Clinical Amyloid Imaging. Semin Nucl Med. 2017;47:31-43.
6. Mathis CA, Mason NS, Lopresti BJ, Klunk WE. Development of Positron Emission Tomography β-Amyloid Plaque Imaging Agents. Semin Nucl Med. 2012;42:423-432.
7. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [18 F]flutemetamol. NeuroImage. 2015;121:184-192.
8. Becker GA, Ichise M, Barthel H, et al. PET Quantification of 18F-Florbetaben Binding to -Amyloid Deposits in Human Brains. J Nucl Med. 2013;54:723-731.
9. Golla SS, Verfaillie SC, Boellaard R, et al. Quantification of [18F]florbetapir: A test–retest tracer kinetic modelling study. J Cereb Blood Flow Metab. June 2018:0271678X18783628.
10. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. NeuroImage. 2008;42:76-86.
11. van Berckel BNM, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. J Nucl Med. 2013;54:1570-1576.
12. Cselényi Z, Farde L. Quantification of Blood Flow-Dependent Component in Estimates of Beta-Amyloid Load Obtained Using Quasi-Steady-State Standardized Uptake Value Ratio. J Cereb Blood Flow Metab. 2015;35:1485-1493.
13. Daerr S, Brendel M, Zach C, et al. Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases. NeuroImage Clin. 2016;14:77-86.
14. Tiepolt S, Hesse S, Patt M, et al. Early [18F]florbetaben and [11C]PiB PET images are a surrogate biomarker of neuronal injury in Alzheimer's disease. Eur J Nucl Med Mol Imaging. 2016;43:1700-1709.
15. Cecchin D, Barthel H, Poggiali D, et al. A new integrated dual time-point amyloid PET/MRI data analysis method. Eur J Nucl Med Mol Imaging. 2017;44:2060-2072.

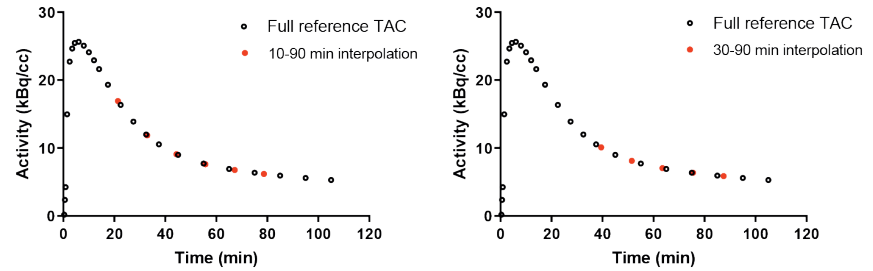
16. Florek L, Tiepolt S, Schroeter ML, et al. Dual Time-Point [18F]Florbetaben PET Delivers Dual Biomarker Information in Mild Cognitive Impairment and Alzheimer's Disease. J Alzheimers Dis JAD. October 2018.
17. Bullich S, Barthel H, Koglin N, et al. Validation of Non-Invasive Tracer Kinetic Analysis of 18 F-Florbetaben PET Using a Dual Time-Window Acquisition Protocol. J Nucl Med. November 2017:jnumed.117.200964.
18. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 Study of the Pittsburgh Compound B Derivative 18F-Flutemetamol in Healthy Volunteers and Patients with Probable Alzheimer Disease. J Nucl Med. 2009;50:1251-1259.
19. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996;4:153-158.
20. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. NeuroImage. 1997;6:279-287.
21. Ichise M, Liow J-S, Lu J-Q, et al. Linearized Reference Tissue Parametric Imaging Methods: Application to [11 C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human Brain. J Cereb Blood Flow Metab. 2003;23:1096-1112.
22. Logan J, Fowler JS, Volkow ND, Wang G-J, Ding Y-S, Alexoff DL. Distribution Volume Ratios without Blood Sampling from Graphical Analysis of PET Data. J Cereb Blood Flow Metab. 1996;16:834-840.
23. Hume SP, Myers R, Bloomfield PM, et al. Quantitation of carbon-11-labeled raclopride in rat striatum using positron emission tomography. Synap N Y N. 1992;12:47-54.
24. Heurling K, Leuzy A, Zimmer ER, Lubberink M, Nordberg A. Imaging β-amyloid using [18F]flutemetamol positron emission tomography: from dosimetry to clinical diagnosis. Eur J Nucl Med Mol Imaging. 2016;43:362-373.
25. Gunn RN, Gunn SR, Cunningham VJ. Positron Emission Tomography Compartmental Models. J Cereb Blood Flow Metab. 2001;21:635-652.
26. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19:716-723.
27. Yaqub M, Boellaard R, Kropholler MA, Lammertsma AA. Optimization algorithms and weighting factors for analysis of dynamic PET studies. Phys Med Biol. 2006;51:4217-4232.
28. Price JC, Klunk WE, Lopresti BJ, et al. Kinetic Modeling of Amyloid Binding in Humans using PET Imaging and Pittsburgh Compound-B. J Cereb Blood Flow Metab. 2005;25:1528-1547.
29. Schuitemaker A, van Berckel BN, Kropholler MA, et al. Evaluation of Methods for Generating Parametric (R)-[11 C]PK11195 Binding Images. J Cereb Blood Flow Metab. 2007;27:1603-1615.

30. Salinas CA, Searle GE, Gunn RN. The simplified reference tissue model: model assumption violations and their impact on binding potential. J Cereb Blood Flow Metab. 2015;35:304-311.
31. Villemagne VL, Ong K, Mulligan RS, et al. Amyloid Imaging with 18F-Florbetaben in Alzheimer Disease and Other Dementias. J Nucl Med. 2011;52:1210-1217.
32. Bremmer JP, van Berckel BNM, Persoon S, et al. Day-to-Day Test–Retest Variability of CBF, CMRO2, and OEF Measurements Using Dynamic 15O PET Studies. Mol Imaging Biol. 2011;13:759-768.

Supplementary materials



Supplementary Figure 1. Interpolation of two different intervals in a reference tissue TAC for [^{18}F]flutemetamol.



Supplementary Figure 2. Interpolation of two different intervals in a reference tissue TAC for [^{18}F]florbetaben.

Supplementary Table 1a. Boundary values of [^{18}F]flutemetamol kinetic parameters

	K_1		k_2		k_3		k_3/k_4		Vt		V_b	
	min	max	min	max	min	max	min	max	min	max	min	max
2T4k_V$_b$	0	1			0.005	1	0 20	20	0	50	0	0.5
FRTM	0	100	0.005	1	0.001	1	0	20				
SRTM	0	100	0.005	1			0	20				

Supplementary Table 1b. Boundary values of [^{18}F]florbetaben kinetic parameters

	K_1		k_2		k_3		k_3/k_4		Vt		V_b	
	min	max	min	max	min	max	min	max	min	max	min	max
2T4k_V$_b$	0	1			0.001	1	0	20	0	50	0	0.5
FRTM	0	100	0.005	1	0.001	1	0	20				
SRTM	0	100	0.005	1			0	20				

4.

Parametric imaging of dual-time window [^{18}F]flutemetamol and [^{18}F]florbetaben scans

Fiona Heeman, Maqsood Yaqub, Janine Hendriks, Ilona Bader, Frederik Barkhof, Juan Domingo Gispert, Bart N.M. van Berckel, Isadora Lopes Alves, Adriaan A. Lammertsma
On behalf of the AMYPAD Consortium

***As published in NeuroImage*, 2021; 234; 117956**

Abstract

Optimal pharmacokinetic models for quantifying amyloid beta (Aβ) burden using both [^{18}F]flutemetamol and [^{18}F]florbetaben scans have previously been identified at a region of interest (ROI) level. The purpose of this study was to determine optimal quantitative methods for parametric analyses of [^{18}F]flutemetamol and [^{18}F]florbetaben scans. Forty-six participants were scanned on a PET/MR scanner using a dual-time window protocol and either [^{18}F]flutemetamol (*N*=24) or [^{18}F]florbetaben (*N*=22). The following parametric approaches were used to derive DVR estimates: reference Logan (RLogan), receptor parametric mapping (RPM), two-step simplified reference tissue model (SRTM2) and multilinear reference tissue models (MRTM0, MRTM1, MRTM2), all with cerebellar grey matter as reference tissue. In addition, a standardized uptake value ratio (SUVR) was calculated for the 90-110 min post injection interval. All parametric images were assessed visually. Regional outcome measures were compared with those from a validated ROI method, i.e. DVR derived using RLogan. Visually, RPM, and SRTM2 performed best across tracers and, in addition to SUVR, provided highest AUC values for differentiating between Aβ-positive vs Aβ-negative scans ([^{18}F]flutemetamol: range AUC=0.96-0.97 [^{18}F]florbetaben: range AUC=0.83-0.85). Outcome parameters of most methods were highly correlated with the reference method (R^2≥0.87), while lowest correlation were observed for MRTM2 (R^2=0.71-0.80). Furthermore, bias was low (≤5%) and independent of underlying amyloid burden for MRTM0 and MRTM1. The optimal parametric method differed per evaluated aspect; however, the best compromise across aspects was found for MRTM0 followed by SRTM2, for both tracers. SRTM2 is the preferred method for parametric imaging because, in addition to its good performance, it has the advantage of providing a measure of relative perfusion (R_1), which is useful for measuring disease progression.

Key words: Amyloid PET, [^{18}F]flutemetamol, [^{18}F]florbetaben, Parametric imaging, PET quantification

Introduction

Early detection of amyloid-beta (Aβ) plaques has become increasingly relevant, in particular with respect to identifying individuals for secondary prevention trials and monitoring disease progression in Alzheimer's disease (AD) (*1*,*2*). These Aβ plaques can be identified *in vivo* with positron emission tomography (PET) scans, using either a dichotomous classification of Aβ burden (*3*) or a more fine-grained quantitative measure of Aβ, which is required for measuring the extent of pathology or quantitatively tracking disease progression (*4*).

For the latter quantitative applications, it has been suggested that dynamic or dual-time window scanning protocols should be used for obtaining the most accurate Aβ estimates (*5*–*7*) and, since amyloid pathology might not follow anatomical boundaries, voxel-wise quantitative analyses might be preferred as they provide more detailed spatial information, independent of predefined regions. In addition, quantitative parametric images allow for voxel-by-voxel analysis between groups or as a function of time within the same subjects. Thus, comprehensive evaluation of parametric methods used for voxel-wise analysis is warranted. However, as different tracers have different kinetics, the optimal parametric method for quantification needs to be established for each Aβ tracer separately. In addition, the ability to obtain a voxel-wise measure of the relative tracer influx rate (R_1) could play a role when deciding which method is most suitable for parametric analysis. This additional parameter can be used as a proxy for (relative) cerebral blood flow (*8*) and may therefore be considered as a measure of disease severity or progression, as first shown by single photon emission computed tomography (SPECT) and [^{15}O]H_2O-PET studies, and more recently replicated with arterial spin labelling (ASL) (*9*–*11*).

For most amyloid tracers, voxel-wise quantitative approaches have been evaluated to a certain extent (*6*,*12*–*14*). However, [^{18}F]flutemetamol and [^{18}F]florbetaben studies were performed with full dynamic acquisitions that did not include the recommended 90-110 min post injection (p.i.) scanning window used for visual assessment and to derive standard update value ratios (SUVR) (*14*,*15*). Further, only Heurling and colleagues evaluated parametric methods for full dynamic [^{18}F]flutemetamol studies by comparing results with those obtained using the gold standard, i.e. full kinetic analysis using an arterial plasma input function (*14*). They showed that best correlations with the gold standard were obtained using reference Logan (RLogan), the 70-90 minutes standardized

uptake value ratio ($SUVR_{70-90}$) and receptor parametric mapping (RPM) (R^2>0.94). With respect to full dynamic [^{18}F]florbetaben studies, the performance of two parametric methods has been evaluated only visually, while regional reference tissue models have been quantitatively compared with the gold standard (*15*). It remains unclear whether the results from full dynamic acquisitions remain valid for a dual-time window protocol, given that the performance of parametric methods could be compromised by the gap in the data. Previous work has shown that in some cases, a gap in the data may lead to sub-optimal curve fitting, which in turn may result in a small bias in the results (*7*). This impact is mediated by the effect of noise and may potentially affect parametric methods in different ways due to differences in their sensitivity to noise. In addition, small interpolation or coregistration errors could also affect performance of parametric methods. These shortened acquisition protocols are especially relevant for fluorine-18 tracers such as [^{18}F]flutemetamol and [^{18}F]florbetaben, considering that routine acquisition of lengthy dynamic scans (110 minutes duration) is not feasible, especially for older participants (*6*,*7*). In addition to increased patient comfort, these protocols can increase patient throughput and result in more cost-efficient use of tracer batch productions compared with full dynamic acquisitions. While still less efficient than standard static acquisitions, a dual-time window can provide an additional parameter (R_1), which can be advantageous for studies focusing on disease progression (*16*–*18*). These characteristics seem to have increased the popularity of dual-time window acquisitions, as can be seen by the adoption of these protocols by large-scale studies such as AMYPAD (*1*).

The present study aimed to evaluate the performance of the most widely used parametric methods for the specific case of dual-time window scans using [^{18}F]flutemetamol and [^{18}F]florbetaben. Performance of these parametric methods was evaluated with respect to three different aspects: 1) visual assessment of parametric images 2) ability to differentiate between Aβ-positive and Aβ-negative scans, and 3) quantitative accuracy and precision with respect to a reference method.

Materials and Methods

Participants

Forty-six cognitively unimpaired subjects were selected from the AMYPAD PNHS (*19*). In short, all subjects underwent neurological and neuropsychological assessment and had at least one PET and one MR scan available. All PET scans were visually assessed by

a trained nuclear medicine physician (BvB) following the formal guidelines for reading [^{18}F]flutemetamol and [^{18}F]florbetaben scans, as defined by the respective manufacturers. Approximately half of the participants (*N*=24) was scanned using [^{18}F]flutemetamol and the other half (*N*=22) using [^{18}F]florbetaben. For both tracers, 50% of the participants was Aβ-positive. All participants provided written informed consent before participating in the study in accordance with the Declaration of Helsinki. The Medical Ethics Review Committee of Amsterdam UMC, location VUmc, approved the study protocol (EudraCT Number: 2018-002277-22).

Image acquisition

All subjects underwent a dynamic PET scan on a Philips Ingenuity TF PET/MR scanner (Philips Medical Systems, Best, The Netherlands), according to a dual-time window protocol (*7*). This scanning protocol consisted of an initial dynamic scan from 0 to 30 minutes p.i. followed by a break of 60 minutes, and then a second scan from 90 to 110 minutes p.i. Prior to each of the scans, a T1-weighted gradient echo pulse MR sequence was acquired for attenuation correction purposes. At the start of the early PET scan, participants received a bolus injection of either [^{18}F]flutemetamol (*N*=24, 184±11 MBq) or [^{18}F]florbetaben (*N*=22, 280±18 MBq), following manufacturer dosage guidelines. Next, scans were reconstructed into 22 frames (6x5, 3x10, 4x60, 2x150, 2x300, and 1x600 s for the first part of the scan, and 4x300 s for the second part) with a matrix size of 128 × 128 × 90 and a voxel size of 2 x 2 x 2 mm with a standard line-of-response–based row-action maximum-likelihood algorithm (LOR-RAMBLA) (*20*) for the brain. All usual corrections, e.g. for MR-based attenuation (MRAC) (*21*), decay, scatter, randoms and dead time were performed. In addition, structural T1-weighted MR images were acquired within, on average, 4.6±3.0 months from the PET scan for [^{18}F]flutemetamol and 4.0±2.5 months for [^{18}F]florbetaben (maximum difference between PET and MR scan was one year), respectively, also on the same Philips Ingenuity TF PET/MR scanner.

Image processing

First, all PET scans were visually checked for between-frame movement. As none showed motion, all were included for further analysis. Next, both structural T1-weighted MR images and early PET images (0-30 min p.i.) were co-registered to their corresponding late PET images (90-110 min p.i.) and visually checked. This was carried out in two steps: first the T1-weighted MR and early MRAC image were co-registered with the late MRAC image and then, the transformation matrix corresponding to the early MRAC was applied to each frame of the early PET image sequence. Following co-registration,

the early and late parts of the dual-time window PET scan were combined into a single file using an in-house developed, MATLAB-based software tool. Subsequently, a PVE-lab-based implementation of SPM8 (*22*) was used to segment the MR image into grey matter, white matter and cerebrospinal fluid (CSF). The reference tissue (cerebellar grey matter) region of interest (ROI) was delineated based on the Hammers atlas (*23*) and time-activity curves (TACs) were extracted.

Parametric analysis

Given that within the AMYPAD PNHS study no arterial input data were acquired, a regional reference tissue method was used as standard to evaluate the performance of all parametric methods. RLogan was chosen as the "regional standard" based on previous studies in which good results in terms of correlation and bias compared with the gold standard (arterial input) had been observed for this method, and stability against variation in confounding factors had also been demonstrated (*8*,*24*).

First, missing data points from the reference tissue TACs were interpolated using the reversible two tissue compartment model (4 rate constants) with additional blood volume fraction parameter (2T4k_V_b) and a typical, tracer-specific plasma input function, based on the interpolation procedure described previously (*7*).

Next, the PPET software tool (*25*) was used to compute either distribution volume ratio (DVR), non-displaceable binding potential (BP_{ND}) or standardized uptake value ratio (SUVR) parametric images based on the following methods: RLogan, receptor parametric mapping (RPM), the voxel-based implementation of the two step simplified reference tissue model (SRTM2), multilinear reference tissue models (MRTM0, MRTM1, MRTM2), and SUVR (calculated for the 90-110 minutes p.i. interval), all with cerebellar grey matter as reference tissue (*26–30*). In addition, parametric relative delivery (R_1) images were generated using RPM and SRTM2. For SRTM2, k_2' was determined across all voxels with a BP_{ND} higher than 0.05 from a first RPM run, while this parameter was omitted in PPET's implementation of RLogan. In line with previous studies, the RLogan linearization start time (t^*) was set to 50 minutes p.i. for both tracers (*8*,*14*,*26*). For both tracers, RPM's basis function (BF) settings were first optimized. The optimization procedure consisted of evaluating a range of BF settings (see supplementary materials, Table S1) and defining the settings that resulted in the highest correlation and the lowest bias (as assessed by R^2 and slope of the regression line) between RPM and regional

standard-derived DVR. To allow for comparability across methods, BP_{ND}+1 values (here referred to as DVR) were computed for methods with BP_{ND} as outcome parameter.

Finally, the following cortical grey matter volumes of interest (VOIs) were superimposed on the parametric images in order to extract regional values for comparison with those estimated using the regional standard: anterior and posterior cingulate gyrus, middle and orbitofrontal gyrus, inferior and superior frontal gyrus, gyrus rectus, pre- and post-central gyrus, superior parietal gyrus, (infero)lateral remainder of the parietal lobe, insula, cuneus, lateral remainder of the occipital lobe, medial and lateral anterior temporal lobe, posterior temporal lobe, superior, middle and inferior temporal gyrus, lingual gyrus, parahippocampal and ambient gyrus, and the fusiform gyrus. Finally, a volume-weighted global cortical average was calculated from all regional data.

Statistical analysis

All statistical analyses were performed in IBM SPSS Statistics for Windows Version 26.0 (IBM Corp. Armonk New York U.S.A) or GraphPad Prism for Windows Version 7.04, (La Jolla California U.S.A.). These analyses can be divided into three groups: 1) assessment of population equivalence between [^{18}F]flutemetamol and [^{18}F]florbetaben scans, 2) visual assessment of parametric images 3) quantitative assessment of the performance of parametric methods compared with the regional standard.

Population equivalence

First, a between tracer comparison of potential differences in age and mini-mental state examination (MMSE) scores of the participants was performed using non-parametric Mann-Whitney *U* tests, while potential differences in the proportion of males and females were compared using a chi-square test. Subsequently, a within tracer sub-group comparison of age and MMSE scores between Aβ-positive and Aβ-negative subjects was performed using the same tests as described above.

Visual assessment of parametric images

Image artifacts were identified first visually and subsequently by calculating the percentage of (extreme) outliers within total grey matter using a threshold that was defined based on expected clinical values: 2.00 for BP_{ND} and 3.00 for DVR and SUVR images (*14*,*15*). Values larger than these thresholds were considered to be outliers. In addition, a more lenient and more stringent threshold were applied to demonstrate the

robustness of the analyses, i.e. 1.80, 2.20 for BP_{ND} images and 2.80, 3.20 for DVR and SUVR images.

Quantitative performance of parametric methods

Performance of each parametric method was first determined by assessing possible differences in global cortical DVR, SUVR and R_1 derived from Aβ-positive and Aβ-negative scans. This was assessed using one-way analysis of variance (ANOVA) and the area under the curve (AUC) from a Receiver Operating Characteristic (ROC) curve which is indicative of how well a parameter is able to differentiate between Aβ-positive and Aβ-negative scans.

Next, linear regression analyses, corrected for age, sex and visual status, were used to assess the correlation (R^2) between binding estimates obtained using each parametric method and the regional standard. In addition, Bland-Altman analyses were used to assess bias compared to the regional standard, the variability of the data based on the 95% Limits of Agreement, and whether proportional bias was present. Any proportional bias was further determined by fitting a regression line through the Bland-Altman plot.

Results

Population equivalence

There were no significant between-tracer differences with respect to the participant's age, proportion of males and females or average MMSE scores (Table 1). Within each tracer group, MMSE scores were higher for Aβ-negative compared with Aβ-positive participants ([^{18}F]flutemetamol: p=0.014, [^{18}F]florbetaben: p=0.061), while there were no differences in age or proportion males and females (Table 1).

Table 1. Subject demographics split by visual Aβ status

	[^{18}F]flutemetamol			[^{18}F]florbetaben		
	All (*N*=24)	**Aβ-negative** (*N*=12)	**Aβ-positive** (*N*=12)	**All** (*N*=22)	**Aβ-negative** (*N*=11)	**Aβ-positive** (*N*=11)
Age	71.7± 6.0	69.9± 5.1	73.5± 6.5	69.4±7.4	68.1± 7.4	70.6±7.5
Females (%)	58.8	75.0**	41.7	63.6	72.7	54.5
MMSE	28.3±1.7	29.2± 1.1*	27.5±1.9	28.9±1.2	29.5± 0.7**	28.4± 1.4

Values depicted as mean ± SD, * p<0.05, ** p<0.10, compared with the Aβ-positive group

Visual assessment of parametric images

Representative parametric BP_{ND}, DVR, SUVR and R_1 images of both Aβ-positive and Aβ-negative scans are presented in Figure 1 for each tracer. Visual inspection of parametric BP_{ND}, DVR and SUVR images showed that, for both tracers, most image artifacts caused by outliers (i.e. speckles) were present for MRTM2 and MRTM1 (Figure 1). Visually, no artifacts were observed in parametric R_1 images, and differences in R_1 derived from Aβ-positive and Aβ-negative scans were small. Furthermore, the quantitative outlier analysis showed that for [^{18}F]flutemetamol, most outliers were present in MRTM2, followed by MRTM1 and SUVR images (range: 0.3-2.8%, supplementary Table S2a). For [^{18}F]florbetaben, most outliers (range: 0.6-4.7%) were present in MRTM2 followed by MRTM1 images (supplementary Table S2b). For both tracers, results showed a very similar pattern using the alternative thresholds (supplementary Tables S2a and S2b).

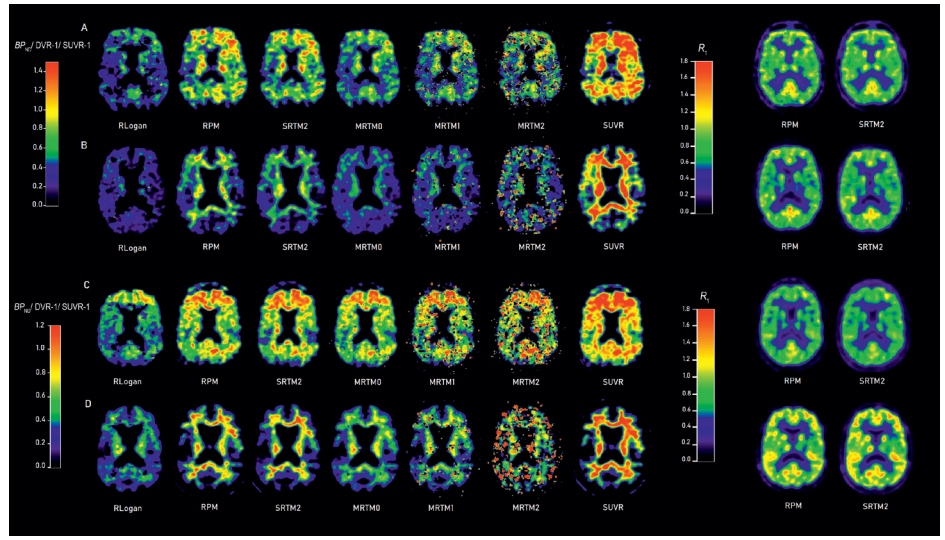


Figure 1. Parametric amyloid and relative perfusion images

Parametric images of each of the methods for **(a)** an Aβ-positive and **(b)** an Aβ-negative [^{18}F]flutemetamol scan, and for **(c)** an Aβ-positive and **(d)** an Aβ-negative [^{18}F]florbetaben scan. *Note: [^{18}F]florbetaben images are shown in colour for comparison, while grey scale is recommended for visual assessment.*

Quantitative performance of parametric methods

For both tracers, all parametric methods showed a higher Aβ burden (as shown by DVR or SUVR values) for the Aβ-positive group ($p<0.05$). Nevertheless, RPM and SRTM2-derived DVR and SUVR yielded highest area under the curve (AUC) values, as shown in Table 2. As expected, for both tracers there were no significant differences in R_1 between Aβ groups and AUC values of RPM and SRTM2-derived R_1 were low (Table 2).

Table 2 AUC: Models' ability to correctly classify Aβ-positive and Aβ-negative scans

	[^{18}F]flutemetamol	[^{18}F]florbetaben
RRLogan DVR	0.94	0.79
RLogan DVR	0.88	0.77
RPM DVR	0.97	0.85
SRTM2 DVR	0.96	0.84
MRTM0 DVR	0.94	0.80
MRTM1 DVR	0.95	0.80
MRTM2 DVR	0.81	0.79
SUVr	0.96	0.83
RPM R_1	0.54	0.45
SRTM2 R_1	0.49	0.45

RRLogan corresponds to the regional implementation of reference Logan (regional standard). Values correspond to the Area Under the Curve (AUC) from ROC analyses

With respect to [^{18}F]flutemetamol, linear regression analyses yielded high correlations ($R^2\geq0.87$) between outcome measures from all methods and regional standard-derived DVR, except for MRTM2 where the correlation was $R^2=0.71$ (Table 3 and Figure 2A). Furthermore, compared with the regional standard, MRTM2-derived DVR showed a constant, average overestimation of 7%, as shown by the Bland-Altman analyses (Figure 3A). On the other hand, SUVR and parametric RLogan-derived DVR showed a bias of 10 and -5% compared with the regional standard, which was proportional to the underlying amyloid burden ($SUVR_{slope}= 0.49$, $p<0.001$, parametric $RLogan_{slope}= -0.20$, $p<0.001$) (Figure 3A). In addition, proportional bias was also observed for RPM and SRTM2 ($RPM_{slope}= 0.23$, $p<0.001$ ($SRTM2_{slope}= 0.15$, $p<0.001$). Furthermore, Bland-Altman analyses showed most variability for MRTM2 and RPM-derived DVR and SUVR (Figure 3A).

With respect to [^{18}F]florbetaben, linear regression analyses yielded high correlations ($R^2\geq0.94$) between outcome measures of all parametric methods and regional standard-derived DVR, except for MRTM2 where the correlation was $R^2=0.80$ (Table 3 and Figure 2B). Furthermore, compared with the regional standard, MRTM2-derived DVR showed a constant, average overestimation of 11%, while SRTM2 showed a constant, average underestimation of 7%, as shown by the Bland-Altman analyses (Figure 3B). On the other hand, SUVR, parametric RLogan and RPM-derived DVR showed a small average bias (<5%), however, this bias was proportional to the underlying amyloid burden ($SUVR_{slope}= 0.31$, $p<0.001$, parametric $RLogan_{slope}= -0.15$, $p<0.001$, $RPM_{slope}= 0.15$, $p<0.001$) (Figure 3B). Finally, Bland-Altman analyses showed most variability for MRTM2-derived DVR and SUVR (Figure 3B).

Table 3. Linear regression analysis

[^{18}F]flutemetamol	Parametric RLogan	RPM	SRTM2	MRTM0	MRTM1	MRTM2	SUVR
R^2	0.95	0.95	0.95	0.99	0.99	0.71	0.87
Slope	0.85	1.14	1.06	0.97	0.98	0.95	1.34
Intercept	0.20	-0.26	-0.18	0.05	0.06	0.15	-0.59
[^{18}F]florbetaben	**Parametric RLogan**	**RPM**	**SRTM2**	**MRTM0**	**MRTM1**	**MRTM2**	**SUVR**
R^2	0.98	0.97	0.95	0.99	0.98	0.80	0.94
Slope	0.87	1.07	0.94	1.02	1.02	0.93	1.22
Intercept	0.11	-0.20	-0.03	-0.09	-0.07	0.20	-0.39

Regional RLogan was used as independent variable, and linear regression analyses were corrected for age, sex and visual read

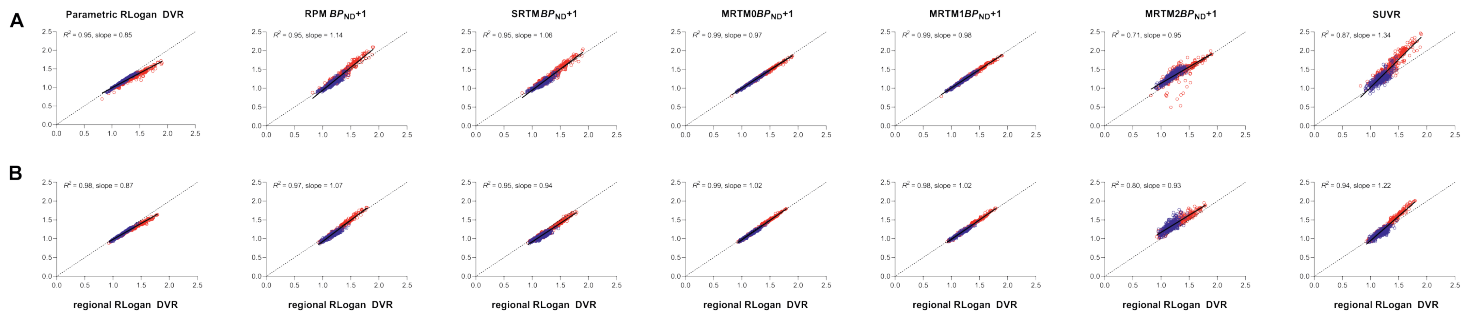


Figure 2. Correlations between outcome measures of all parametric methods and the regional standard for [^{18}F]flutemetamol and [^{18}F]florbetaben

Correlations shown for **(a)** [^{18}F]flutemetamol and **(b)** [^{18}F]florbetaben, with regional RLogan DVR as regional standard for all regions of interest (ROIs). ROIs are colour coded based on the Aβ status of the scan they belong to, i.e. Aβ-positive = red, Aβ-negative = blue. The dashed line corresponds to the line of identity and the solid line corresponds to the regression line.

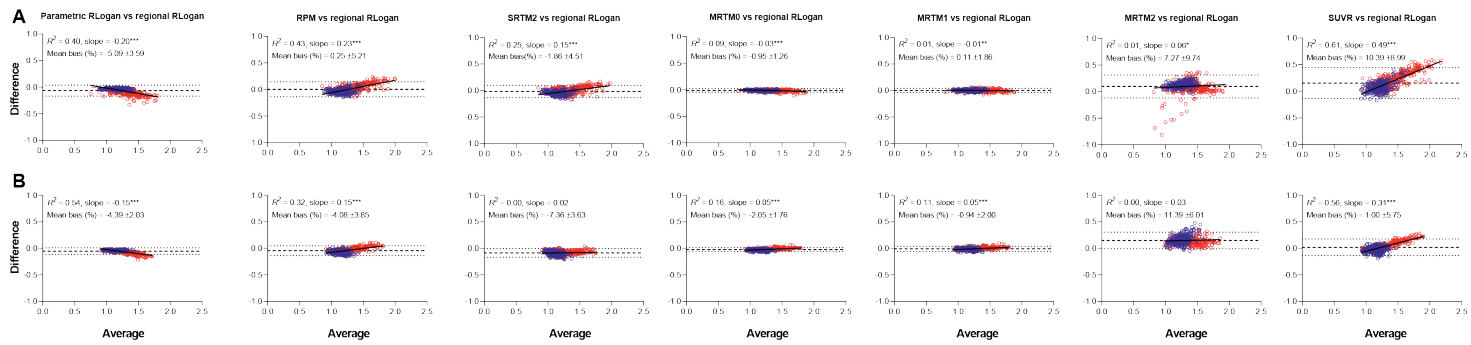


Figure 3. Bland-Altman plots showing the agreement between outcome measures of all parametric methods and the regional standard for [^{18}F]flutemetamol and [^{18}F]florbetaben

Bland-Altman plots for **(a)** [^{18}F]flutemetamol and **(b)** [^{18}F]florbetaben, showing the absolute difference between outcome measure of each parametric method and regional RLogan DVR for all regions of interest (ROIs). ROIs are colour-coded based on the Aβ status of the scan they belong to, i.e. Aβ-positive = red, Aβ-negative = blue. The dashed horizontal line corresponds to the average bias, the dotted horizontal lines correspond to the upper and lower limit of the 95% Limits of Agreement for all scans and the solid line to the linear regression of the BA data points. *** p<0.001, ** p<0.01, * p<0.05.

Discussion

In this study, various parametric methods for voxel-wise analysis of dual-time window [^{18}F]flutemetamol and [^{18}F]florbetaben studies were evaluated, considering the following three different aspects: 1) visual assessment of parametric images 2) ability to differentiate between Aβ-positive and Aβ-negative scans 3) quantitative accuracy and precision. All evaluated methods could differentiate between Aβ-positive and Aβ-negative scans based on DVR or SUVR measures, and they all showed high correlations with regional standard-derived DVR, except for MRTM2. However, the quality of parametric images and the level of bias compared with the regional standard varied substantially across methods. Hence, these findings demonstrate that the most appropriate parametric method depends on the research or clinical question to be addressed.

For both tracers, high correlations (R^2≥0.87) were observed between DVR derived using the regional standard and the outcome measure from all parametric methods, except for MRTM2, which had relatively lower correlations ([^{18}F]flutemetamol: R^2=0.71 and [^{18}F]florbetaben: R^2=0.80) and showed a constant, average DVR overestimation of 7% for [^{18}F]flutemetamol and 11% for [^{18}F]florbetaben compared with the regional standard. It should be noted, that for interpreting the bias or accuracy of the results, it is important to take into account whether the bias is proportional to the underlying Aβ burden. In contrast to a constant bias, proportional bias cannot easily be accounted for in longitudinal studies and thereby pose challenges, in particular when aiming to measure small changes over time. In that regard, while small (max. 5%) on average (except for [^{18}F]flutemetamol SUVR), the bias observed with the other methods was proportional to the underlying Aβ burden, especially in the case of SUVR, RPM and RLogan. For SRTM2 specifically, a small proportional bias was observed for [^{18}F]flutemetamol, while a constant, average underestimation of 7% was seen for [^{18}F]florbetaben.

The amyloid dependent bias, as reported above, could be explained by a variety of factors. One possible explanation could be the assumption of single-tissue compartment kinetics in both the target and reference tissues, which is assumed for RPM and SRTM2, but not for MRTM implementations (*27–30*). In fact, it has been reported that for these tracers the reference tissues' kinetics are better described by a 2TC model (*15*,*24*), possibly explaining the amyloid dependent bias for these methods, as demonstrated by Salinas and colleagues (*32*). This assumption may be particularly problematic for regions with high Aβ burden, explaining the increase in bias as illustrated by Figure

3. With respect to RLogan, the voxel-wise method showed marked differences with its regional implementation counterpart. Since all other aspects (e.g. linearization start time, method of interpolation for missing data points and omission of k_2') were identical between both implementations, these differences should primarily be related to a general impact of noise on parametric imaging methods. In particular, it has been shown that underestimation of the amyloid burden with RLogan-derived DVR is more pronounced at higher noise levels and for regions with higher binding (*33*). Therefore, the parametric RLogan bias dependency on underlying Aβ burden is plausibly a result of the interaction between higher noise levels and a higher Aβ burden (*33*).

In general, the present results are well in line with previous amyloid PET studies, although for some methods, such as RPM and SUVR in case of [^{18}F]flutemetamol, the bias was larger than previously reported (*14*). For example, previous work has demonstrated an overestimation of SUVR compared with BP_{ND} or DVR obtained using plasma input modelling (the gold standard) (*5*,*13–15*,*34*,*35*). Since the reference method used in the present study (regional RLogan) is known to underestimate true binding (*5*,*13*,*14*,*35*), this could explain the larger levels of bias observed in SUVR and RPM, compared with those reported by Heurling and colleagues (*14*). However, given that the regional implementation of RLogan is less sensitive to noise compared with its parametric counterpart, and the Aβ burden covered by the CU participants is expected to be slightly lower than what is typically observed in AD patients, limited impact on the results was expected. With respect to both quantitative and visual performance of the various MRTM implementations, considerable between-tracer differences have been reported previously (*12–15*). These differences in performance may be explained by the application (or lack) of pre- or post-reconstruction smoothing filters and differences in scanner resolution (*12–15*). Finally, small differences in the performance of parametric methods compared with previous reports can also be attributed to specific choices in processing and analysis pipelines, such as differences in SUVR uptake times, both starting and boundary values of fitting parameters, reference tissue selection and/or the inclusion of subcortical regions (*14*,*15*). Of note, for a few data points, large differences were observed between MRTM2 and the regional standard in case of [^{18}F]flutemetamol (Figure 3A). All these data points belonged to a single subject and no large differences were observed for any of the other methods. Unfortunately, there is no obvious reason why data for this subject were different only for MRTM2.

Overall, the aim of this study was to evaluate different parametric methods on three different aspects. Regarding the first: *visual assessment of parametric images*, few image artifacts were observed for RLogan, RPM, SRTM2, SUVR and MRTM0, facilitating visual assessment. Furthermore, with respect to the tracer distribution, RPM and SRTM2, followed by MRTM0 images, appear most suitable. On the other hand, SUVR images showed high uptake in grey as well as white matter regions, increasing the risk of false positives, while RLogan images showed clearest underestimation of the Aβ burden, increasing the risk of false negatives, especially for [^{18}F]flutemetamol (Figure 1). Therefore, RPM and SRTM2 appear to be the methods of choice for visual assessment, followed by MRTM0. The second aspect was *the ability to differentiate between Aβ-positive and Aβ-negative scans*. For both tracers, all methods could detect significant differences between Aβ-positive and Aβ-negative scans based on DVR or SUVR measures and were therefore considered suitable for this purpose. However, RPM, SRTM2 and SUVR provided highest AUC values, making these the preferred methods for detecting more subtle differences in Aβ burden, followed by MRTM0 and MRTM1. Regarding the third aspect, *quantitative accuracy and precision*, methods should show a high correlation and low bias (independent of underlying Aβ burden) compared with the regional standard and low measurement variability. With respect to the first characteristic, all methods, except from MRTM2, would be considered appropriate, as all of them showed excellent correlations with the regional standard ($R^2 \geq 0.87$). Regarding the second, only MRTM0 and MRTM1 showed both minimal and constant bias (<5%), while all other methods showed a larger and/or proportional bias to a certain degree. Finally, lowest variability, as shown by the Limits of Agreement of the BA analyses (Figure 3), was observed for MRTM0 and MRTM1, followed by parametric RLogan, SRTM2 and RPM, in case of [^{18}F]florbetaben. Therefore, MRTM0 and MRTM1 appear to be the most suitable methods with respect to accuracy and precision.

Of the preferred methods from the first and second aspect, RPM and SRTM2 have the additional benefit of providing a measure of relative perfusion (R_1) compared with the MRTM implementations. In the present study, R_1 yielded only low AUC values for differentiating between Aβ-positive and Aβ-negative scans. However, this finding was expected, as a cognitively unimpaired population was studied and changes in neuronal function and CBF tend to manifest at a later stage in the disease than Aβ accumulation (*36*).

A limitation of this work is the lack of a direct comparison with the gold standard, plasma input modelling, to determine quantitative accuracy and precision, which could

be done by future studies. No large differences would be expected, as several studies have shown that reference tissue approaches are adequate for quantification compared with the (plasma input) gold standard (*13*,*14*,*37*). In addition, good results have been reported previously for RLogan in terms of correlation and bias compared with the gold standard and its robustness against confounding factors (*8*,*24*). Furthermore, the performance of the parametric methods was evaluated using cognitively unimpaired subjects only. To cover a range of Aβ burden, 50% Aβ-positive and 50% Aβ-negative subjects were selected. Nonetheless, the upper end of this range may be slightly lower than what is typically observed in AD dementia patients. It should be noted that, as the subjects scanned according to a dual-time window protocol will unlikely include patients with severe cognitive impairment (given that this is not the target population for clinical trials or studies measuring disease progression), the present range of Aβ burden was considered appropriate for the goal of this study.

Conclusions

The preferred parametric methods for voxel-based amyloid quantification of [^{18}F]flutemetamol and [^{18}F]florbetaben dual-time window studies differed per evaluated aspect, but were relatively comparable between tracers. Compared with the reference standard method, regional RLogan, the best compromise across aspects was found for MRTM0, followed closely by SRTM2, which has the advantage of also providing R_1. Given the current interest in R_1, SRTM2 would be the preferred parametric method.

Acknowledgements

The authors would like to thank Ronald Boellaard for his technical assistance, Lyduine Collij for coordinating collection of the scans and the staff of the department of Radiology and Nuclear Medicine of the Amsterdam UMC, location VUmc for skilful acquisition of the scans.

References

1. Farrar G, Barkhof F, Pla S, et al. AMYPAD: A EUROPEAN PUBLIC-PRIVATE PARTNERSHIP TO INVESTIGATE THE VALUE OF β-AMYLOID BRAIN SCANS AS A DIAGNOSTIC AND THERAPEUTIC MARKER FOR ALZHEIMER'S DISEASE. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2017;13:P420.
2. Collij LE, Heeman F, Salvadó G, et al. Multi-tracer model for staging cortical amyloid deposition using PET imaging. Neurology. July 2020:10.1212/WNL.0000000000010256.
3. Mallik A, Drzezga A, Minoshima S. Clinical Amyloid Imaging. Semin Nucl Med. 2017;47:31-43.
4. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. J Nucl Med. 2017;58:1019-1024.
5. Berckel BNM van, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. J Nucl Med. 2013;54:1570-1576.
6. Bullich S, Barthel H, Koglin N, et al. Validation of Noninvasive Tracer Kinetic Analysis of 18F-Florbetaben PET Using a Dual-Time-Window Acquisition Protocol. J Nucl Med. 2018;59:1104-1110.
7. Heeman F, Yaqub M, Lopes Alves I, et al. Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. EJNMMI Research. 2019;9:32.
8. Heeman F, Yaqub M, Alves IL, et al. Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies: J Cereb Blood Flow Metab. 2021;41:579-589.
9. Collij LE, Heeman F, Kuijer JPA, et al. Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer Disease. Radiology. 2016;281:865-875.
10. Binnewijzend MAA, Benedictus MR, Kuijer JPA, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. Eur Radiol. 2016;26:506-514.
11. Jagust WJ, Eberling JL, Reed BR, Mathis CA, Budinger TF. Clinical Studies of Cerebral Blood Flow in Alzheimer's Disease. Annals of the New York Academy of Sciences. 1997;826:254-262.
12. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008;42:76-86.
13. Verfaillie SC, Golla SS, Timmers T, et al. Repeatability of parametric methods for [18F] florbetapir imaging in Alzheimer's disease and healthy controls: A test–retest study. J Cereb Blood Flow Metab. 2021;41.
14. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18)F]flutemetamol. Neuroimage. 2015;121:184-192.

15. Becker GA, Ichise M, Barthel H, et al. PET quantification of 18F-florbetaben binding to β-amyloid deposits in human brains. J Nucl Med. 2013;54:723-731.
16. Tiepolt S, Hesse S, Patt M, et al. Early [18F]florbetaben and [11C]PiB PET images are a surrogate biomarker of neuronal injury in Alzheimer's disease. European Journal of Nuclear Medicine and Molecular Imaging. 2016;43:1700-1709.
17. Son SH, Kang K, Ko P-W, et al. Early-Phase 18F-Florbetaben PET as an Alternative Modality for 18F-FDG PET. Clinical Nuclear Medicine. 2020;45:e8.
18. Daerr S, Brendel M, Zach C, et al. Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases. Neuroimage Clin. 2016;14:77-86.
19. Lopes Alves I, Collij LE, Altomare D, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. Alzheimer's Dement. April 2020.
20. Hu Z, Wang W, Gualtieri EE, et al. An LOR-based fully-3D PET image reconstruction using a blob-basis function. In: 2007 IEEE Nuclear Science Symposium Conference Record. Vol 6. ; 2007:4415-4418.
21. Hu Z, Renisch S, Schweizer B, et al. MR-based attenuation correction for whole-body PET/MR system. In: IEEE Nuclear Science Symposuim Medical Imaging Conference. ; 2010:2119-2122.
22. Rask T, Dyrby T, Comerci M, et al. PVElab: Software for correction of functional images for partial volume errors. Neuroimage. 2004;22.
23. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp. 2003;19:224-247.
24. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 Study of the Pittsburgh Compound B Derivative 18F-Flutemetamol in Healthy Volunteers and Patients with Probable Alzheimer Disease. Journal of Nuclear Medicine. 2009;50:1251-1259.
25. Boellaard R, Yaqub M, Lubberink M, Lammertsma A. PPET: A software tool for kinetic and parametric analyses of dynamic PET studies. NeuroImage. 2006;Supplement 2:T62.
26. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. J Cereb Blood Flow Metab. 1996;16:834-840.
27. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. NeuroImage. 1997;6:279-287.
28. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996;4:153-158.
29. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. J Cereb Blood Flow Metab. 2002;22:1440-1452.

30. Ichise M, Liow J-S, Lu J-Q, et al. Linearized Reference Tissue Parametric Imaging Methods: Application to [11C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human Brain. Journal of Cerebral Blood Flow & Metabolism. 2003;23:1096-1112.
31. Hinkle DE, Wiersma William, Jurs SG. Applied statistics for the behavioral sciences. 5th ed. Boston, Mass.: Houghton Mifflin ; 2003.
32. Salinas CA, Searle GE, Gunn RN. The simplified reference tissue model: model assumption violations and their impact on binding potential. J Cereb Blood Flow Metab. 2015;35:304-311.
33. Slifstein M, Laruelle M. Effects of Statistical Noise on Graphic Analysis of PET Neuroreceptor Studies. J Nucl Med. 2000;41:2083-2088.
34. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. Validation of the Semiquantitative Static SUVR Method for 18F-AV45 PET by Pharmacokinetic Modeling with an Arterial Input Function. J Nucl Med. 2017;58:1483-1489.
35. Heeman F, Hendriks J, Lopes Alves I, et al. [11C]PiB amyloid quantification: effect of reference region selection. EJNMMI Research. 2020;10:123.
36. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010;9:119-128.
37. Price JC, Klunk WE, Lopresti BJ, et al. Kinetic Modeling of Amyloid Binding in Humans using PET Imaging and Pittsburgh Compound-B. J Cereb Blood Flow Metab. 2005;25:1528-1547.

Supplementary methods

Table S1 shows the basis function settings that were evaluated for both tracers. Start parameters were derived from previously published [^{11}C]PiB data (Settings A) and a smaller range from minimum to maximum was preferred, because of the higher processing efficiency (*1*).

Table S1. Basis function settings

Settings	Min	Max	Steps
A	0.010	0.3	40
B	0.010	0.3	100
C	0.010	0.4	40
D	0.010	0.2	40
E	0.005	0.2	40
F	0.005	0.3	40
G	0.005	0.4	40
H	0.050	0.2	40
I	0.050	0.3	40
J	0.050	0.4	40

Supplementary results

The optimisation procedure of RPM's basis function settings showed that, for both tracers, increasing the number of steps from 40 to 100 did not improve results. Furthermore, 0.01 was the optimal minimum, as increasing or decreasing this number negatively affected correlation with the regional standard. Adjusting the maximum value had a negligible effect on correlation, but a small effect on bias. The optimal basis setting were the following, for [^{18}F]flutemetamol: min=0.01, max=0.30, steps=40 and for [^{18}F]florbetaben: min=0.01, max=0.20, steps=40.

Artifacts parametric images

Table S2a. Outliers in parametric [^{18}F]flutemetamol images

		A	B	C
RLogan	Grey	0.0	0.0	0.0
RPM	Grey	0.1	0.0	0.0
SRTM2	Grey	0.0	0.0	0.0
MRTM0	Grey	0.0	0.0	0.0
MRTM1	Grey	0.7	0.6	0.6
MRTM2	Grey	3.5	2.8	2.7
SUVr	Grey	0.8	0.3	0.1

Outliers (%) for 3 different outlier thresholds. A= 1.8 and 2.8, B=2.0 and 3.0, C=2.2 and 3.2, for BP_{ND} and DVR/SUVR, respectively

Table S2b. Outliers in parametric [^{18}F]florbetaben images

		A	B	C
RLogan	gm	0.0	0.0	0.0
RPM	gm	0.0	0.0	0.0
SRTM2	gm	0.0	0.0	0.0
MRTM0	gm	0.0	0.0	0.0
MRTM1	gm	0.7	0.6	0.6
MRTM2	gm	5.6	4.7	4.3
SUVR	gm	0.0	0.0	0.0

Outliers (%) for 3 different outlier thresholds. A= 1.8 and 2.8, B=2.0 and 3.0, C=2.2 and 3.2, for BP_{ND} and DVR/SUVR images, respectively

References

1. Yaqub M, Tolboom N, Boellaard R, van Berckel BNM, van Tilburg EW, Luurtsema G, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008 Aug 1;42(1):76–86.

5.

Test-retest variability of relative tracer delivery rate as measured by [^{11}C]PiB

Fiona Heeman, Janine Hendriks, Isadora Lopes Alves, Nelleke Tolboom, Bart N.M. van Berckel, Maqsood Yaqub, Adriaan A. Lammertsma

As published in Molecular Imaging and Biology, 2021; 23(3): 335-339.

Abstract

Purpose: Moderate-to-high correlations have been reported between the [^{11}C]PiB PET derived relative tracer delivery rate R_1 and relative CBF as measured using [^{15}O]H_2O PET, supporting its use as a proxy of relative CBF. As longitudinal PET studies become more common for measuring treatment efficacy or disease progression, it is important to know the intrinsic variability of R_1. The purpose of the present study was to determine this through a retrospective data analysis.

Procedures: Test-retest data belonging to twelve participants, who underwent two 90 min [^{11}C]PiB PET scans, were retrospectively included. The voxel-based implementation of the two-step simplified reference tissue model with cerebellar grey matter as reference tissue was used to compute R_1 images. Next, test-retest variability was calculated and test and retest R_1 measures were compared using Linear Mixed Effect Models and a Bland-Altman analysis.

Results: Test-retest variability was low across regions (max. 5.8%), and test and retest measures showed high, significant correlations (R^2=0.92, slope=0.98) and a negligible bias (0.69±3.07%).

Conclusions: In conclusion, the high precision of [^{11}C]PiB R_1 suggests suitable applicability for cross-sectional and longitudinal studies.

Keywords: [^{11}C]PiB, Alzheimer's disease, cerebral blood flow, relative tracer delivery, test-retest variability

Background

Cerebral blood flow (CBF) is known to decline with age, and elderly individuals (75-80 years) may present with reductions in CBF of up to 25% compared with young adults (±25 years) (*1*,*2*). In the context of Alzheimer's disease (AD), additional reductions in CBF have been reported in several cortical brain regions such as the frontal, parietal and temporal cortices with both absolute reductions as well as relative to cerebellar grey matter reference tissue (*3*,*4*). This pattern of CBF reductions is considered characteristic of AD pathology and may therefore be used as proxy for measuring disease severity or progression (*5*,*6*). The gold standard technique for measuring CBF is [^{15}O]H_2O positron emission tomography (PET) (*7*), but MR-based techniques such as arterial spin labelling (ASL) have also been introduced (*5*). More recently, several studies have evaluated whether a valid proxy of cerebral perfusion can be obtained from the early frames of dynamic scans using currently available PET tracers (e.g. for measuring amyloid-β or tau burden) (*3*,*8*). The relative influx rate (R_1=K_1/K_1') which can be calculated from these early frames, is an indirect measure of relative CBF as it is also affected by the extraction fraction (K_1=E·CBF). In particular, for the amyloid tracer [^{11}C]PiB, Chen and colleagues have reported that in cortical regions, relative tracer delivery R_1 showed moderate-to-high correlations with relative CBF measured using dynamic [^{15}O]H_2O PET (with a range of ρ=0.68-0.84, p<0.001, across cortical regions and a value of ρ=0.82, p<0.001, for the global cortical region), thereby indicating that [^{11}C]PiB could be used for dual-biomarker imaging (*9*). Furthermore, Bilgel and colleagues demonstrated that in a longitudinal setting with an average follow-up duration of 2.5 years, rates of change as measured with R_1 showed low-to-moderate correlations with changes in [^{15}O]H_2O PET CBF (median r =0.42 across cortical regions and a value of r =0.65 for the global cortical region). In addition, rates of change as measured with R_1 required the smallest sample-size for detecting group-wise differences (27% reduction compared with [^{15}O]H_2O PET), suggesting this proxy could be suitable for tracking long-term longitudinal changes in CBF (*8*). As longitudinal PET studies in AD become more common for measuring treatment efficacy or disease progression, it is important to know the intrinsic variability of R_1 in order to determine what magnitude of change in R_1 signifies an actual change. Therefore, the purpose of the present retrospective analysis of a previously reported test-retest (TRT) [^{11}C]PiB study was to assess the precision of [^{11}C]PiB R_1.

Materials & Methods

Subjects

Data from twelve participants belonging to a TRT study conducted within the Amsterdam UMC, location VUmc, were reanalysed as the original study only reported TRT variability for the non-displaceable binding potential (*10*). This dataset consisted of five cognitively unimpaired (CU) subjects, one patient with mild cognitive impairment (MCI) and six with AD dementia, which in the present study was used to examine TRT variability for R_1 (*10*). Before enrolment, written informed consent was obtained from all individual participants included in the study and the Medical Ethics Review Committee of the Amsterdam UMC, location VUmc, had approved the study.

Image acquisition

All subjects underwent a structural T1-weighted MR scan on a 1.5T Siemens Sonata scanner and, within one week, two 90 minutes dynamic [^{11}C]PiB PET scans (test and retest) on a Siemens ECAT EXACT HR+ scanner (*10*). Each dynamic scan consisted of 23 consecutive time frames (1×15, 3×5, 3×10, 2×30, 3×60, 2×150, 2×300, 7×600 s). All participants received an intravenous injection of, on average 332±70 and 376±43 MBq, for test and retest scans, respectively. More specifically, CU participants received 353±26 and 355±37 MBq, the MCI patient 138 and 368, MBq and AD patients 342±66 and 393±47 MBq for test and retest scans, respectively.

Image processing

Structural T1-weighted MR images were co-registered to their corresponding PET image segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) using PVE-lab software (*11*). Next, volumes of interest (VOIs) were delineated based on the Hammers atlas and a reference tissue time–activity curve (TAC) of the cerebellar grey matter was extracted (*12*,*13*).

Parametric analysis

The PPET software tool (*14*) with the voxel-based implementation of the two-step simplified reference tissue model (SRTM2), as validated for [^{11}C]PiB, and cerebellar grey matter as reference tissue were used to compute relative tracer delivery (R_1) images (*15–17*). For SRTM2, k_2' was determined across all voxels with a BP_{ND} higher than 0.05 by taking the median k_2' from a first run using receptor parametric mapping (RPM) (*18*). Regional R_1 values were obtained by superimposing the following grey matter VOIs on

the parametric images: medial and lateral anterior temporal lobe, posterior temporal lobe, superior, middle and inferior temporal gyrus, fusiform gyrus, parahippocampal and ambient gyrus, anterior and posterior cingulate gyrus, middle and orbitofrontal gyrus, gyrus rectus, inferior and superior frontal gyrus, pre- and post-central gyrus, superior parietal gyrus and the (infero)lateral remainder of the parietal lobe and a global cortical composite region (i.e. volume-weighted average across all target regions).

Statistical analysis

Statistical analyses were performed in R (version 4.0.3; R Foundation for Statistical Computing, Vienna, Austria). First, global R_1 values were compared between CU and AD groups using a non-parametric Mann-Whitney *U* test, separately for test and retest scans. Next, TRT variability was calculated for regional and global cortical R_1 values according to Equation 1, where T represents the estimate of R_1 measured during test, and *R* the one measured during retest.

$$TrT\ variability\ (\%) = \frac{|T - R|}{0.5 \cdot |T + R|} \cdot 100 \qquad (1)$$

In addition, a correlation analysis was used to assess the relationship between TRT variability and regional volume. Furthermore, to assess the relationship between test and retest R_1 measures, Linear Mixed Effect Models (LME) were fitted and correlation coefficients were calculated using the nlme and MuMIn packages, respectively (*19*,*20*). Visual read (amyloid-β positive or negative) was used as a covariate and the analysis accounted for the within-subject correlation between regions. Finally, a Bland-Altman analysis was used to assess potential bias between test and retest R_1 using the Blandr package (*21*,*22*).

Results

Participant characteristics are shown in Table 1. Relative tracer delivery measures (R_1) are reported in Table 2, with a significantly lower global R_1 in AD patients compared with CU participants, for both test and retest scans ($p < 0.01$).

Table 1. Subject demographics

	CU (*N*=5)	MCI (*N*=1)	AD (*N*=6)
Age	64.6 ±6.4	71.0	61.0 ±3.0
Females	60%	100%	17%
VR positive	20%	0%	100%
MMSE	29.8 ±0.4	28.0	20.7 ±2.0

VR: Visual read, MMSE: Mini-Mental State Examination
Values are depicted as Mean±SD, unless indicated otherwise

Table 2. Relative tracer delivery values by diagnostic group

	SRTM2-derived R_1	
Diagnostic groups	**Test**	**Retest**
CU (*N*=5)	0.93 ± 0.04	0.91 ± 0.03
MCI (*N*=1)	0.91	0.91
AD (*N*=6)	0.82 ± 0.04	0.82 ± 0.03

Values are depicted as Mean±SD

Regional and global cortical TRT variability values are presented in Table 3. TRT variability for the global cortical composite was low (1.70%), while the range of regional TRT variability showed slightly higher values (range: 1.52-5.78%). Furthermore, there was a trend effect towards smaller TRT variability for larger regions (R^2=0.14, p=0.09).

Table 3. Regional test-retest variability (%) of R_1

Region	SRTM2-derived R_1
Global Cortex	1.70
Anterior temporal lobe medial part	2.77
Anterior temporal lobe lateral part	2.48
Parahippocampal and ambient gyri	3.02
Superior temporal gyrus	2.68
Middle and inferior temporal gyri	2.87
Fusiform gyrus	2.26
Insula	2.06
Lateral remainder of occipital lobe	2.08
Gyrus cinguli anterior part	1.52
Gyrus cinguli posterior part	3.44
Middle frontal gyrus	2.13
Posterior temporal lobe	2.11
Inferolateral remainder of parietal lobe	2.31
Precentral gyrus	1.71
Gyrus rectus	5.78

Table 3. Regional test-retest variability (%) of R_1 (continued)

Region	SRTM2-derived R_1
Orbitofrontal gyri	3.06
Inferior frontal gyrus	2.00
Superior frontal gyrus	2.08
Postcentral gyrus	1.79
Superior parietal gyrus	1.59
Lingual gyrus	2.60
Cuneus	1.76

LME analyses showed that test and retest R_1 values were strongly correlated and that the slope was not significantly different from 1 (R^2=0.92, slope=0.98 C.I. [0.94-1.01], p<0.001). Furthermore, amyloid status as measured by visual read did not have a significant effect on this relationship. Finally, Bland-Altman analysis showed a negligible bias (0.69±3.07%) between test and retest R_1 (Figure 1). All analyses were also carried out using RPM-derived R_1 which resulted in essentially identical results (data not shown).

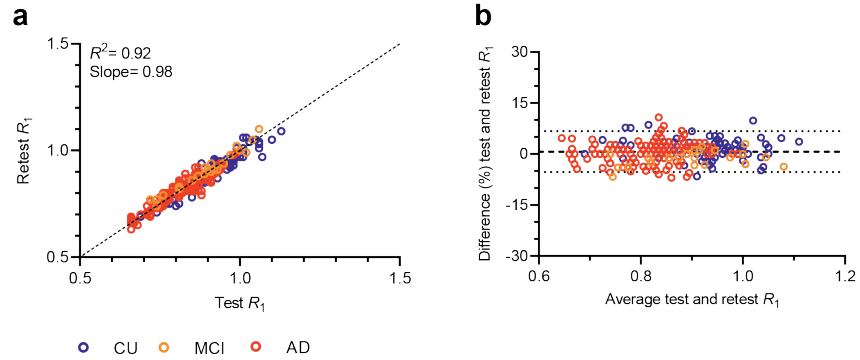


Figure 1. Relationship between SRTM2-derived test and retest R_1
(a) The correlation between R_1 test and retest measures, with R_2 and slope parameters corresponding to the LME analysis and **(b)** a Bland-Altman plot, which indicates the bias between the two measures.

Discussion

The present study assessed precision of the R_1 parameter, a measure of relative tracer delivery, through a retrospective analysis of a previously reported [^{11}C]PiB test-retest study (*10*). Low test-retest variability was observed for SRTM2-derived R_1 and this was true for regions of different sizes.

Differences in R_1 between diagnostic groups were as expected, with lower average R_1 values in AD patients compared with CU participants. This finding is in agreement with existing literature where decreases in (relative) perfusion related to AD pathology have been reported for both R_1 and [^{15}O]H_2O PET studies (*3*,*23*,*24*). Furthermore, by incorporating these two participant groups, the present study covered the entire range of R_1 values that would be expected in clinical studies across the AD spectrum.

Excellent TRT variability was observed for the global cortical composite (1.70%) and only a slightly poorer TRT variability for some of the smaller regions (max 5.8%). These findings were supported by the results of the LME analysis which showed a high correlation between test and retest R_1 measures (R^2=0.92) and a slope that was close to identity. As expected, the results indicate that smaller TRT variability was associated with larger regions. This finding suggests that studies should consider looking at relatively larger regions with PET when their aim is to detect small (<5%) changes.

Despite distinct kinetics, the present findings were also comparable with results from a [^{18}F]florbetapir study that assessed TRT variability of SRTM derived-R_1 in a very similar population in terms of age and diagnosis (max. TRT variability of 6%) (*25*). Comparing [^{11}C]PiB R_1 TRT variability with TRT variability of absolute perfusion as measured with the gold standard, [^{15}O]H_2O PET (*26*), shows that the present results are slightly better, indicating that R_1 is a more precise measure likely due to the fact that it is a relative measure. Furthermore, given that R_1 is an indirect measure of perfusion, variation in the extraction fraction may compensate day-to-day fluctuations in flow to maintain constant delivery. On the other hand, this also means that alterations in extraction fraction may bias R_1, as opposed to the gold standard [^{15}O]H_2O PET which provides a direct measurement of CBF. Unfortunately, to date, no studies have reported TRT variability of relative perfusion as measured with [^{15}O]H_2O PET with cerebellar grey matter reference tissue for a direct comparison. Nonetheless, a study by Bilgel and colleagues compared long term variability of CBF proxies (i.e. [^{11}C]PiB PET-derived R_1

and early frame standardized uptake value ratios) to that of [^{15}O]H_2O PET and reported that, across these three measures, the highest longitudinal stability was obtained with R_1. The present study demonstrates that R_1 also has low short term variability. Therefore, R_1 could be considered as a valid alternative to measuring longitudinal changes in CBF, thereby circumventing the need for a separate [^{15}O]H_2O PET scan. Unfortunately, it remains unclear whether changes across a lifetime are comparable between R_1 and rCBF as measured by [^{15}O]H_2O PET. One study reports that, in an elderly population (77±8 years old), the yearly percentage change in R_1 was lower (-0.28%) than that in rCBF (-0.41%) as measured by [^{15}O]H_2O PET, although it was not reported whether this difference was significant (*8*). Nevertheless, a smaller change in R_1, especially in elderly subjects, may be related to an increased extraction fraction (*27*). However, longitudinal studies of R_1 in a younger population are needed to confirm whether the same difference is present earlier in life.

Furthermore, using a relative parameter to measure CBF such as R_1 essentially assumes that there are no CBF changes in the reference tissue (R_1=K_1/K_1'). In this regard, it should be noted that differences in whole cerebellum CBF (a commonly used reference tissue) have been reported when comparing AD patients and age-matched controls (*28*). In contrast, such differences have not been demonstrated for cerebellar cortex CBF by studies using a similar design in terms of technique and participants (*24*,*29–31*). This suggests that careful interpretation is required when comparing longitudinal R_1 measurement between AD patients and controls or that alternative reference tissues, unaffected by CBF changes, should be considered. Yet, further research is required to understand whether such changes in cerebellar CBF also occur in early AD stages.

Conclusion

Relative tracer delivery rate R_1 of [^{11}C]PiB showed high global and regional precision in participants covering the AD spectrum. Therefore, [^{11}C]PiB R_1 appears to be a stable parameter for measuring cross-sectional differences and longitudinal changes in relative CBF.

References

1. Pantano P, Baron JC, Lebrun-Grandié P, Duquesnoy N, Bousser MG, Comar D. Regional cerebral blood flow and oxygen consumption in human aging. Stroke. 1984;15:635-641.
2. Leenders KL, Perani D, Lammertsma AA, et al. Cerebral blood flow, blood volume and oxygen utilization. Normal values and effect of age. Brain. 1990;113 (Pt 1):27-47.
3. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. 18F-FDG PET, the early phases and the delivery rate of 18F-AV45 PET as proxies of cerebral blood flow in Alzheimer's disease: Validation against 15O-H2O PET. Alzheimer's & Dementia. August 2019.
4. Fukuyama H, Ogawa M, Yamauchi H, et al. Altered Cerebral Energy Metabolism in Alzheimer's Disease: A PET Study. J Nucl Med. 1994;35:1-6.
5. Collij LE, Heeman F, Kuijer JPA, et al. Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer Disease. Radiology. 2016;281:865-875.
6. Wierenga CE, Hays CC, Zlatar ZZ. Cerebral Blood Flow Measured by Arterial Spin Labeling MRI as a Preclinical Marker of Alzheimer's Disease. J Alzheimers Dis. 2014;42:S411-S419.
7. Jagust WJ, Eberling JL, Reed BR, Mathis CA, Budinger TF. Clinical Studies of Cerebral Blood Flow in Alzheimer's Disease. Annals of the New York Academy of Sciences. 1997;826:254-262.
8. Bilgel M, Beason-Held L, An Y, Zhou Y, Wong DF, Resnick SM. Longitudinal evaluation of surrogates of regional cerebral blood flow computed from dynamic amyloid PET imaging. J Cereb Blood Flow Metab. 2020;40:288-297.
9. Chen YJ, Rosario BL, Mowrey W, et al. Relative 11C-PiB Delivery as a Proxy of Relative CBF: Quantitative Evaluation Using Single-Session 15O-Water and 11C-PiB PET. J Nucl Med. 2015;56:1199-1205.
10. Tolboom N, Yaqub M, Boellaard R, et al. Test-retest variability of quantitative [11C]PiB studies in Alzheimer's disease. Eur J Nucl Med Mol Imaging. 2009;36:1629-1638.
11. Rask T, Dyrby T, Comerci M, et al. PVElab: Software for correction of functional images for partial volume errors. Neuroimage. 2004;22.
12. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp. 2003;19:224-247.
13. Heeman F, Hendriks J, Lopes Alves I, et al. [11C]PiB amyloid quantification: effect of reference region selection. EJNMMI Research. 2020;10:123.
14. Boellaard R, Yaqub M, Lubberink M, Lammertsma A. PPET: A software tool for kinetic and parametric analyses of dynamic PET studies. NeuroImage. 2006;Supplement 2:T62.
15. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. J Cereb Blood Flow Metab. 2002;22:1440-1452.

16. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008;42:76-86.
17. Peretti DE, Vállez García D, Reesink FE, et al. Relative cerebral flow from dynamic PiB scans as an alternative for FDG scans in Alzheimer's disease PET studies. PLoS One. 2019;14:e0211000.
18. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. NeuroImage. 1997;6:279-287.
19. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. https://CRAN.R-project.org/package=nlme.
20. Barton K. MuMIn: Multi-Model Inference. https://CRAN.R-project.org/package=MuMIn.
21. Datta D. blandr: a Bland-Altman Method Comparison package for R. Zenodo; 2017.
22. Martin Bland J, Altman DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet. 1986;327:307-310.
23. Deo AK, Borson S, Link JM, et al. Activity of P-glycoprotein, a β-amyloid Transporter at the Blood-Brain Barrier, is Compromised in Patients with Mild Alzheimer's Disease. J Nucl Med. 2014;55:1106-1111.
24. Ishii K, Sasaki M, Yamaji S, Sakamoto S, Kitagaki H, Mori E. Paradoxical hippocampus perfusion in mild-to-moderate Alzheimer's disease. J Nucl Med. 1998;39:293-298.
25. Golla SS, Verfaillie SC, Boellaard R, et al. Quantification of [18F]florbetapir: A test–retest tracer kinetic modelling study. J Cereb Blood Flow Metab. June 2018:0271678X18783628.
26. Bremmer JP, van Berckel BNM, Persoon S, et al. Day-to-day test-retest variability of CBF, CMRO2, and OEF measurements using dynamic 15O PET studies. Mol Imaging Biol. 2011;13:759-768.
27. van Assema DME, Lubberink M, Boellaard R, et al. P-Glycoprotein Function at the Blood–Brain Barrier: Effects of Age and Gender. Mol Imaging Biol. 2012;14:771-776.
28. Kuwabara Y, Ichiya Y, Otsuka M, Masuda K, Ichimiya A, Fujishima M. Cerebrovascular responsiveness to hypercapnia in Alzheimer's dementia and vascular dementia of the Binswanger type. Stroke. 1992;23:594-598.
29. Ishii K, Sasaki M, Matsui M, et al. A diagnostic method for suspected Alzheimer's disease using H(2)15O positron emission tomography perfusion Z score. Neuroradiology. 2000;42:787-794.
30. Cohen RM, Andreason PJ, Doudet DJ, Carson RE, Sunderland T. Opiate receptor avidity and cerebral blood flow in Alzheimer's disease. Journal of the Neurological Sciences. 1997;148:171-180.
31. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. Validation of the Semiquantitative Static SUVR Method for 18F-AV45 PET by Pharmacokinetic Modeling with an Arterial Input Function. J Nucl Med. 2017;58:1483-1489.

PART II

Impact of confounding factors

6.

Simulating the effect of cerebral blood flow changes on regional quantification of [^{18}F]flutemetamol and [^{18}F]florbetaben studies

Fiona Heeman, Maqsood Yaqub, Isadora Lopes Alves, Kerstin Heurling, Santiago Bullich, Juan Domingo Gispert, Ronald Boellaard, Adriaan A. Lammertsma
On behalf of the AMYPAD Consortium

***As published in Journal of Cerebral Blood Flow & Metabolism*, 2021; 41(3): 579-589**

Abstract

Global and regional changes in cerebral blood flow (CBF) can result in biased quantitative estimates of amyloid load by PET imaging. Therefore, the current simulation study assessed effects of these changes on amyloid quantification using a reference tissue approach for [^{18}F]flutemetamol and [^{18}F]florbetaben. Previously validated pharmacokinetic rate constants were used to simulate time-activity curves (TACs) corresponding to full dynamic and dual-time window acquisition protocols. CBF changes were simulated by varying the tracer delivery (K_1) from +25 to -25%. The standardized uptake value ratio (SUVr) was computed and TACs were fitted using reference Logan (RLogan) and the simplified reference tissue model (SRTM) to obtain the relative delivery rate (R_1) and volume of distribution ratio (DVR). RLogan was least affected by CBF changes (χ^2=583 p<0.001, χ^2=81 p<0.001, for [^{18}F]flutemetamol and [^{18}F]florbetaben, respectively) and the extent of CBF sensitivity generally increased for higher levels of amyloid. Further, SRTM derived R_1 changes correlated well with simulated CBF changes (R^2>0.95) and SUVr's sensitivity to CBF changes improved for later uptake-times, with the exception of [^{18}F]flutemetamol cortical changes. In conclusion, RLogan is the preferred method for amyloid quantification of [^{18}F]flutemetamol and [^{18}F]florbetaben studies and SRTM could be additionally used for obtaining a CBF proxy.

Key words: Alzheimer's disease, amyloid PET, cerebral blood flow, quantification, simulation study

Introduction

Amyloid-beta accumulation (Aβ) in the brain is one of the hallmarks of Alzheimer's disease (AD). It can be visualised and quantified using positron emission tomography (PET) and both static or dynamic scanning protocols can be used (*1*,*2*). The static protocol has the advantage of a short scan duration together with relatively simple processing and analytical steps, while the dynamic protocol provides higher accuracy at the cost of a much longer scan, more complex processing and advanced kinetic analysis. The semi-quantitative parameter obtained from a static scan, the standardized uptake value ratio (SUVr), depends on post-injection starting time and duration of the acquisition, and may be affected by changes in cerebral blood flow (CBF) as demonstrated for [^{11}C]PiB (*3*–*6*). In contrast, the non-displaceable binding potential (BP_{ND}), derived from a dynamic PET scan, is less sensitive to noise, more robust against CBF changes, and for [^{11}C]PiB it has been shown to be the parameter of choice when measuring longitudinal changes in amyloid burden (*5*,*6*). This characteristic may be especially important in situations where changes in CBF can occur, such as when measuring disease progression or treatment response in clinical trials. As a compromise between the protocols mentioned above, dynamic data acquisitions from a dual-time window protocol have gained attention, in which data are acquired separately for early and late phases of tracer uptake to reduce overall scanning time, maintain high quantitative accuracy and provide tracer delivery information (*7*,*8*).

Cerebral blood flow declines with age and differs per brain region (*9*). Compared with young adults (25 years), elderly (late 70's) can present with up to 25% CBF reductions, with an average annual CBF decline in grey matter of approximately 0.5% (*9*,*10*). On top of this global decline, day-to-day whole brain CBF fluctuations of around 30% have been reported in test-retest studies, however this percentage was assumed to consist of both physiological as well as measurement error (*11*). Furthermore, drugs may also exert an effect on CBF (*12*,*13*), thus potentially compromising the measurement of amyloid changes associated to pharmacological interventions. In the context of AD, additional relative CBF changes may be present, as focal reductions in CBF have been observed in several brain regions (*14-16*). A study on non-demented older adults scanned with [^{11}C]PiB showed that subjects with elevated [^{11}C]PiB signals experienced greater relative CBF variations, mainly in those regions showing increased amyloid deposition (*17*). Given that CBF differs per region, declines with age, and that additional regional CBF changes occur during the course of Alzheimer's disease, it is important to assess

whether and, if so, to which extent these changes affect quantification of amyloid load. Furthermore, the additional effects of acquisition start-time and duration on quantitative accuracy need to be understood, due to its relevance in clinical and research practice (*3*).

The effect of CBF changes on amyloid load quantification has been assessed previously for [^{11}C]PiB, [^{18}F]florbetaben and semi-quantitative measures of [^{18}F]florbetapir (*5,7,18*). These studies showed that large cortical CBF reductions resulted in a maximum change of ±10% in SUVr. For large global CBF reductions, a larger bias was observed for [^{11}C]PiB as compared with [^{18}F]florbetaben and [^{18}F]florbetapir (*5,7,18*). These between tracer differences are to be expected due to differences in tracer kinetics and corresponding equilibrium times. However, the effect of CBF changes on quantification of [^{18}F]flutemetamol scans remains unknown, just as the potential effects of such changes on novel dual-time window protocols. Therefore, the present simulation study aimed to assess the effects of regional (i.e. target and reference tissues) and global CBF changes on quantitative amyloid measures derived from static, dynamic and dual-time window scanning protocols using reference tissue approaches (*8*), thereby focusing on [^{18}F]flutemetamol and [^{18}F]florbetaben, the two tracers used within the AMYPAD consortium. A second aim was to assess whether the relative delivery rate (R_1) of both ligands could be used to accurately monitor changes in CBF (*16,19*).

Materials and Methods

Subjects and PET data

Previously reported clinical PET data from six [^{18}F]flutemetamol and twenty [^{18}F]florbetaben subjects (both control and AD subjects), consisting of regional time-activity curves (TACs) and whole blood and/or plasma input curves, were used for this simulation study (*20,21*). All participants provided written informed consent in accordance with the Declaration of Helsinki. The Ethical Committee of the University Hospitals Leuven approved the study protocol for the [^{18}F]flutemetamol study (EudraCT 2007-000784-19, Registered 8 February 2007). For [^{18}F]florbetaben, the local Institutional Review Board of University of Leipzig, the National Radiation Safety Committee, and the German Federal Institute for Drugs and Medical Devices approved the study protocol (EudraCT 2006-003882-15, Registered 2006). The first group received a bolus injection (<40 seconds) of 181±5 MBq [^{18}F]flutemetamol and were scanned on a Siemens HiRez Biograph 16 PET/CT scanner (*3,20*). The second group received an injection of 300±60

MBq [^{18}F]florbetaben (over 90 seconds) and were scanned on a ECAT HR+ Siemens/CTI scanner (*21*). As described previously, whole blood curves and metabolite corrected plasma input curves were available for [^{18}F]flutemetamol, and metabolite corrected plasma input curves and discrete whole blood samples for [^{18}F]florbetaben (*8*). Subsequently, regional TACs were fitted using the reversible two-tissue compartment model (four rate constants) with additional blood volume fraction parameter (2T4k_V_b) to obtain pharmacokinetic rate constants for the target (composite cortical region consisting of anterior and posterior cingulate, frontal, parietal, and lateral and medialtemporal cortex) and reference tissue (grey matter cerebellum) (*8*).

TAC simulations

Based on the pharmacokinetic rate constants estimated from the clinical data mentioned above, realistic target and reference tissue TACs of 130 minutes duration were simulated using the 2T4k_V_b model (see Figure 1 and Table 1 for the rate constants) (*8*). Target tissue TACs were simulated for a clinically observed range of BP_{ND} values (captured by five simulated BP_{ND} values, Figure 1), and noise (0, 1 and 2%) corresponding to regions of interest (ROI) of various sizes was added to the target tissue TACs (*22*). This resulted in a total of 50 TACs per noise level (identical TACs in case of 0% noise) and for each simulated BP_{ND} (see Figure 1 for a flowchart of the method). No noise was added to reference tissue TACs due to its relatively large volume.

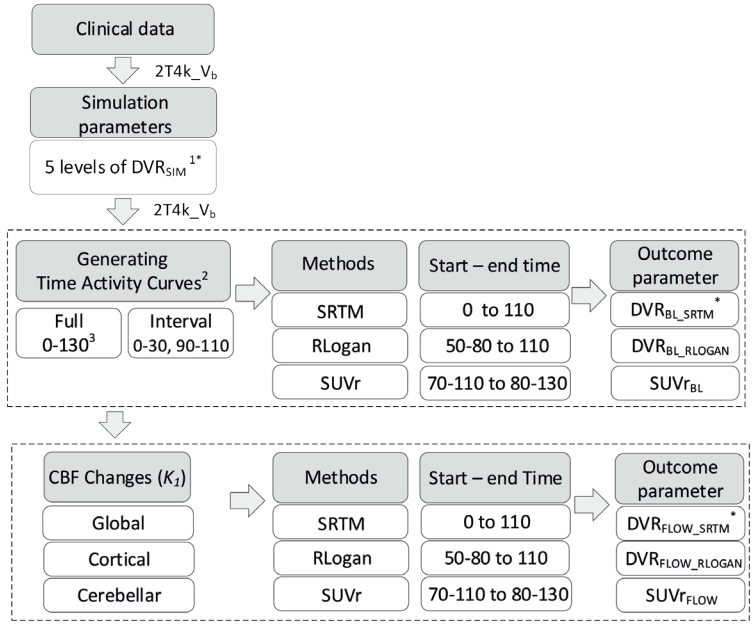


Figure 1. Schematic overview of the applied methods. 1. [^{18}F]flutemetamol DVR_{SIM} (simulated volume of distribution ratio) ranged from 1.022 – 1.778, [^{18}F]florbetaben DVR_{SIM} ranged from 1.026 – 2.051. 2. All steps were repeated for 1 and 2% noise. 3. All values are minutes post injection (p.i.). *Here DVR refers to BP_{ND}+1. Abbreviations: V_b: blood volume, BL: baseline (no cerebral blood flow change), SRTM: simplified reference tissue model, RLOGAN: reference Logan, SUVr: standardised uptake value ratio, FLOW: with changes in cerebral blood flow.

Table 1. Rate constants used for CBF simulations

[^{18}F]flutemetamol	Global		Local cortical		Local cerebellar	
	Target	Reference	Target	Reference	Target	Reference
K_1	0.186-0.310	0.240-0.400	0.186-0.310	0.320*	0.248*	0.240-0.400
k_2	0.060-0.100	0.077- 0.129	0.060-0.010	0.103*	0.080*	0.077- 0.129
k_3	0.008-0.028	0.018	0.008-0.028	0.018	0.008-0.028	0.018
k_4	0.020	0.050	0.020	0.050	0.020	0.050
[^{18}F]florbetaben						
K_1	0.170-0.283	0.188-0.313	0.170-0.283	0.250*	0.226*	0.188-0.313
k_2	0.052- 0.087	0.057-0.095	0.052- 0.087	0.076*	0.069*	0.057-0.095
k_3	0.010-0.030	0.007	0.010-0.030	0.007	0.010-0.030	0.007
k_4	0.010	0.007	0.010	0.007	0.010	0.007

*Baseline parameter, corresponding to a 0% CBF change. Units: K_1 in ml * g^{-1} * min^{-1} and k_2 in min^{-1}. For both tracers, the blood volume fraction parameter (V_b) was set to 0.05.

Simulating CBF changes

Given that the rate constant for tracer delivery (K_1) can be considered a proxy for blood flow as long as the extraction remains constant ($K_1 = Flow * Extraction$), changes in CBF were simulated by varying both K_1 and tissue clearance (k_2), in order to maintain the non-displaceable volume of distribution constant ($V_{ND} = K_1/k_2$) (*23*). Both increases and decreases were simulated (from +25 to -25%), in three different scenarios (see Table 1 and Figure 2 for the simulation parameters and a schematic overview of the simulations):

1. Global CBF changes: equal variations in K_1 and k_2 for target and reference tissues.
2. Cortical CBF changes: variations in K_1 and k_2 only for the target tissue, reference tissue parameters were kept constant.
3. Cerebellar CBF changes: variations in K_1 and k_2 only for the reference tissue, target tissue parameters were kept constant.

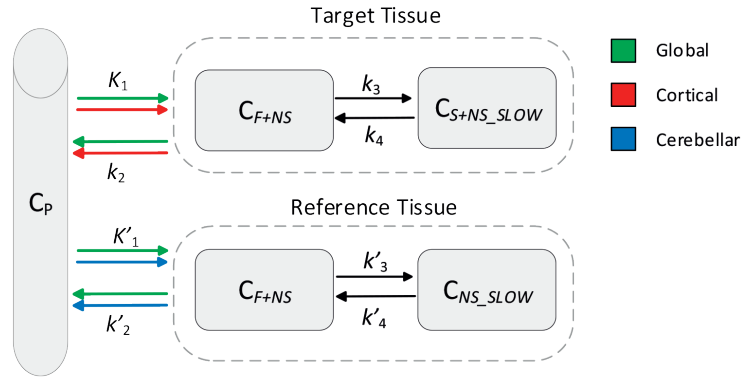


Figure 2. Schematic overview of the simulated CBF changes. C_p: tracer plasma radioactivity concentration, C_{F+NS}: free and non-specifically bound concentration in tissue, C_{S+NS_SLOW}: specifically bound and slow non-specifically bound concentration in tissue, all units Bq*ml^{-1}. Units of the rate constants: K_1 in ml * g^{-1} * min^{-1} and k_2, k_3, k_4 in min^{-1}.

Dual-time window TACs

In addition to the full 130 minutes TACs, dual-time window TACs (here called 'interval TACs') were created, by removing data points from target and reference tissue TACs according to the dual-time window protocol (*8*). In short, the first 110 minutes were used (corresponding to a dynamic scanning protocol) and data points between the early (0-30 minutes post injection (p.i.)) and late acquisition phase (90-110 minutes p.i.) were deleted.

Estimating parameters of interest

SUVr was calculated from the 130 minutes full TACs, for starting times (t^*) ranging from 70 to 110 minutes p.i. (interval t^*: 3x 5min, 4x2.5min, 3x5min), and durations of 10, 15 and 20 minutes. Next, 110 minutes of the full TACs (corresponding to a dynamic scanning protocol) and interval TACs were fitted using reference Logan (RLogan) to estimate the volume of distribution ratio (DVR) (*24*). The implementation of this model did not require fixing k_2' (as per equation 7) (*24*). The linearization start times (t^*) evaluated ranged from 50 to 80 minutes p.i. (in steps of 10 minutes) and interpolation of interval target tissue TACs was performed using cubic interpolation, as the method is routinely used for parametric imaging. Finally, 110 minutes of the full TACs and the interval TACs were fitted using the simplified reference tissue model (SRTM) to estimate R_1, k_2 and BP_{ND} (*25*). Given that SRTM requires a continuous input to fit the target tissue TACs, missing data points of the reference tissue interval TACs were interpolated using the 2T4k_V_b model together with a typical, tracer specific input function as validated previously (*8*). In addition, for both tracers, boundary values (optimized for full, noiseless TACs) were set for all kinetic parameters to prevent physiologically implausible results (Supplementary Figure 1). Optimal boundaries were defined based on the simulated parameter range and lower boundaries were fine-tuned (based on parameter histograms) in case of k_2. This procedure selected the boundary value that resulted in least bias (calculated from the simulated BP_{ND}+1) across CBF scenarios and amyloid binding levels.

Statistical analysis

For all analyses (i.e. all amyloid levels, starting times, durations, scanning protocols and for TACs both with and without noise), parameter estimates obtained in the absence of CBF variations were used as baseline parameters ($SUVr_{\_BL}$, R_{1_BL}, DVR_{BL_SRTM} and DVR_{BL_RLOGAN}, Table 1). Next, percentage change as a result of CBF variations was calculated relative to each baseline for $SUVr_{FLOW}$, R_{1_FLOW}, DVR_{FLOW_SRTM} and RLogan DVR_{FLOW_RLOGAN}:

$$\Delta PAR_{FLOW}\,(\%) = \frac{PAR_{FLOW} - PAR_{BL}}{PAR_{BL}} \cdot 100\% \qquad (1)$$

Where PAR corresponds to the parameter of interest (i.e. SUVr, DVR or R_1). These parameters were then used to assess the sensitivity of the methods to variations in CBF. This was done for different levels of amyloid load, different scanning protocols, noise levels and uptake times.

Non-parametric Kruskal-Wallis tests were used to assess differences in sensitivity to CBF variations between methods and simulated BP_{ND} values (*26*). In addition, the maximum percentage change due to CBF variations was compared between methods and for the whole spectrum of simulated BP_{ND} values. Furthermore, to assess the relationship between ΔR_{1_FLOW} and simulated ΔCBF, the coefficient of determination (R^2) was calculated based on Pearson's correlation coefficient. Next, differences in sensitivity were compared between full and interval TACs fitted with SRTM and RLogan using the post-hoc Mann-Whitney U test and boxplot analyses, where outliers were defined as points outside the whiskers (created using the Tukey method (*27*)). Subsequently, effects of noise on sensitivity to CBF changes and its relationship with all parameters was evaluated based on boxplot analyses. Finally, using Kruskal-Wallis tests, effects of altering starting times (t^*) on CBF sensitivity was evaluated for RLogan and altering starting times and acquisition duration on CBF sensitivity for SUVr.

Results

Evaluation of outcome parameter(s)

The resulting boundary values for [^{18}F]flutemetamol were: R_1: lower boundary (LB)=1e-6 upper boundary (UB)=2.00, k_2: LB=0.018 UB=2.00, BP_{ND}: LB=-0.50 UB=5.00. For [^{18}F]florbetaben, boundary values were: R_1: LB=1e-6 UB=2.00, k_2: LB=0.028 UB=2.00, BP_{ND}: LB=-0.50 UB=5.00.

Sensitivity to CBF changes – comparison between methods

[^{18}F]flutemetamol

Overall, DVR_{FLOW_RLOGAN} was least sensitive to changes in CBF (χ^2=583, p<0.001). With respect to global CBF changes, all models showed less than 6% change. More specifically, DVR_{FLOW_SRTM} showed least overall sensitivity to global CBF changes (χ^2=72, p<0.001), while DVR_{FLOW_RLOGAN} showed the smallest maximum change (3.7%) followed by $SUVr_{FLOW}$ (5.1%) and DVR_{FLOW_SRTM} (max. 5.8%). For cortical and cerebellar changes, DVR_{FLOW_RLOGAN} was least sensitive to CBF changes (χ^2=995 p<0.001, maximum regional change: 2.0% compared with DVR_{FLOW_SRTM} 13.4% and $SUVr_{FLOW}$ 4.8%), as shown in Figure 3. Sensitivity to simulated CBF changes increased for higher simulated levels of amyloid load for global (DVR_{FLOW_SRTM}: χ^2=280, p<0.001; DVR_{FLOW_RLOGAN}: χ^2=104, p<0.001; $SUVr_{FLOW}$: χ^2=239, p<0.001), cortical (DVR_{FLOW_SRTM}: χ^2=63, p<0.001;

DVR_{FLOW_RLOGAN} : χ^2=118, p<0.001; $SUVr_{FLOW}$: χ^2=142, p <0.001) and cerebellar changes (DVR_{FLOW_SRTM} : χ^2 =161, p <0.001; : χ^2 =92, p<0.001; $SUVr_{FLOW}$: no significant difference).

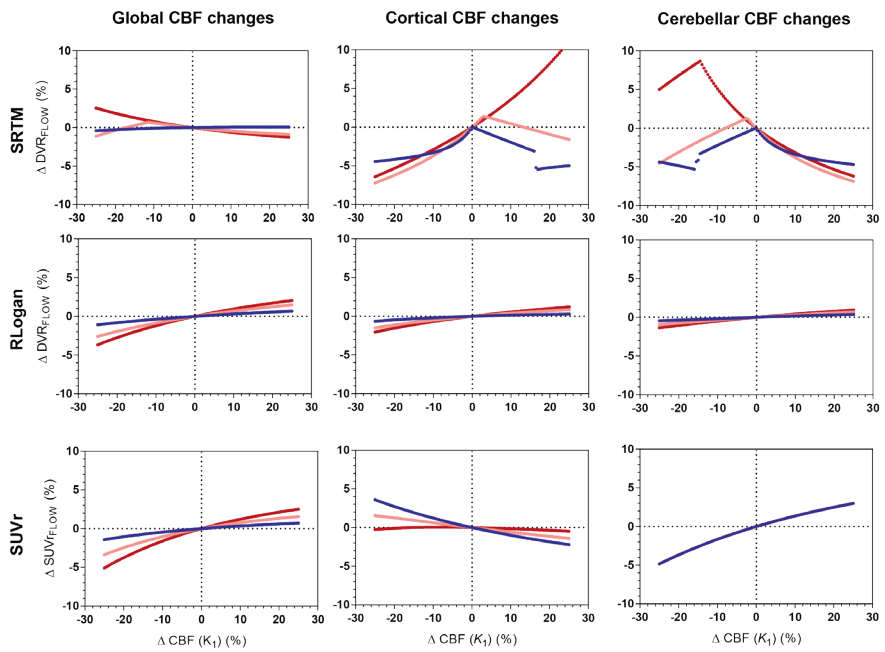


Figure 3. [^{18}F]flutemetamol: sensitivity to CBF changes across methods for three levels of amyloid load. Low amyloid (blue): DVR=1.022, intermediate amyloid (pink): DVR=1.400, high amyloid (red): DVR=1.778. Four values (between 10-10.6%) for SRTM cortical CBF changes high amyloid are not shown. RLogan linearization time was 50-110 minutes p.i., SUVr uptake time was 90-110 minutes p.i.

[^{18}F]florbetaben

Overall, DVR_{FLOW_RLOGAN} was least sensitive to CBF changes compared with the other methods (χ^2 =81, p<0.001), in particular for cerebellar changes (χ^2 =457, p<0.001), and showed smallest maximum change (DVR_{FLOW_RLOGAN} 1.2% DVR_{FLOW_SRTM} 6.1%, $SUVr_{FLOW}$: 2.4%). With respect to global CBF changes, all methods showed a very similar sensitivity pattern with CBF decreases resulting in lower and CBF increases in higher changes (maximum change: DVR_{FLOW_RLOGAN} :5.2%, DVR_{FLOW_SRTM} : 7.5% and $SUVr_{FLOW}$:6.1 %), as shown in Figure 4. With respect to cortical CBF changes, $SUVr_{FLOW}$ was least sensitive (χ^2 =68. p<0.001) and showed the smallest maximum change ($SUVr_{FLOW}$

: 3.7%, DVR_{FLOW_RLOGAN} : 4.7% and DVR_{FLOW_SRTM} : 7.3%). Furthermore, in general the extent of CBF sensitivity increased for higher levels of simulated amyloid load, for global (DVR_{FLOW_SRTM} : χ^2=290, p<0.001; DVR_{FLOW_RLOGAN} : χ^2=171, p<0.001; $SUVr_{FLOW}$: χ^2=274, p<0.001), cortical (DVR_{FLOW_SRTM} : χ^2=131, p<0.001; DVR_{FLOW_RLOGAN} : χ^2=105, p<0.001; SUVr: χ^2=184, p<0.001) and cerebellar changes (DVR_{FLOW_SRTM} : χ^2=13, p=0.011: DVR_{FLOW_RLOGAN} : χ^2=201, p<0.001; $SUVr_{FLOW}$: no significant difference).

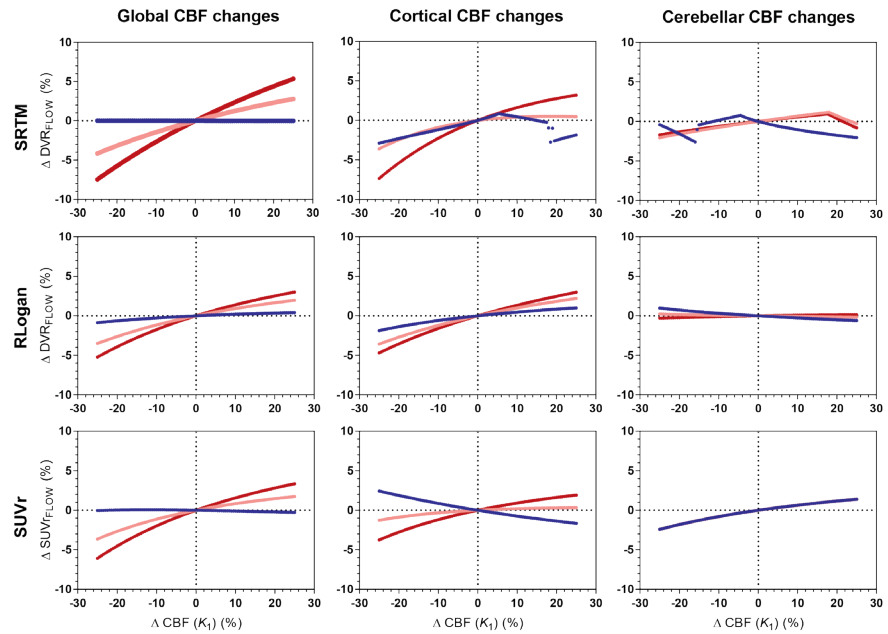


Figure 4. [^{18}F]florbetaben: sensitivity to CBF changes across methods for three levels of amyloid load. Low amyloid (blue): DVR=1.026, intermediate amyloid (pink): DVR=1.538, high amyloid (red): DVR=2.051. RLogan linearization time was 50-110 minutes p.i., SUVr uptake time 90-110 minutes p.i.

Relative delivery and CBF changes

For [^{18}F]flutemetamol, correlations between simulated ΔCBF and ΔR_{1_FLOW} were high for all amyloid levels both for cortical (R^2=0.98 to 1.00, p<0.001) as well as cerebellar CBF changes (R^2=0.97 to 1.00, p<0.001). For [^{18}F]florbetaben, high correlations were also observed between simulated ΔCBF and ΔR_{1_FLOW} across amyloid levels both for cortical (R^2=0.98 to

0.99, p<0.001) and cerebellar CBF changes (R^2=0.95 to 0.99, p<0.001). Very similar results were obtained for R_1 estimates derived from interval TACs (data not shown).

Sensitivity to CBF changes - Full vs. interval TACs

For [^{18}F]florbetaben, full TACs were found to be less sensitive to CBF changes than interval TACs (p<0.001) while for [^{18}F]flutemetamol, the opposite effect was found (p<0.001). Post-hoc tests revealed that for [^{18}F]florbetaben this effect was present for both models, while for [^{18}F]flutemetamol the effect was only present for SRTM. Furthermore, for both tracers, most outliers were present for interval compared to full TACs (outliers interval TACs: [^{18}F]flutemetamol: 91.67%, [^{18}F]florbetaben: 66.67%). These outliers were only observed for TACs fitted with SRTM, predominantly corresponding to low simulated amyloid levels (Figure 5). For [^{18}F]florbetaben, interval TAC-derived DVR_{FLOW_SRTM} was more sensitive to cerebellar CBF changes compared to full TAC-derived DVR_{FLOW_SRTM}, in case of intermediate (Figure 5) to high (data not shown) amyloid levels.

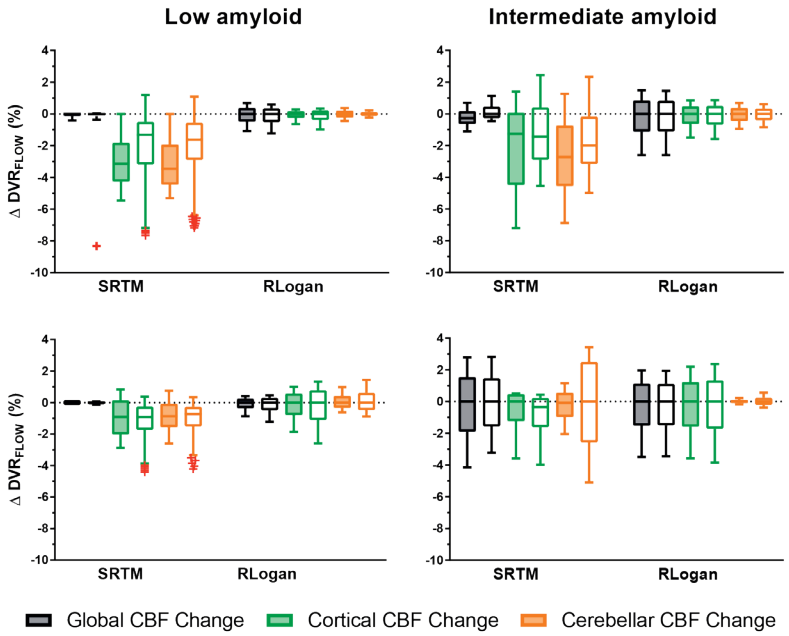


Figure 5. Comparison of full and interval TACs and their sensitivity to CBF changes. Coloured boxes represent full- and uncoloured boxes interval TACs. Upper row [^{18}F]flutemetamol, bottom row [^{18}F]florbetaben. Note: whiskers were defined according to the Tukey method and outliers are depicted as red crosses.

Sensitivity to CBF changes – The effect of noise

Overall, effects of noise on the sensitivity of the different methods to CBF changes were minimal for full TACs (maximum difference: 1 percentage point (p.p.)), see Supplementary Figure 2). Furthermore, for interval TACs, more outliers and a larger sensitivity were observed at higher compared with lower noise levels.

Sensitivity to CBF changes - acquisition start time and duration

Overall, later uptake times showed a decreased sensitivity to CBF changes for both tracers ([^{18}F]flutemetamol: χ^2=54, p<0.001, [^{18}F]florbetaben: χ^2=127, p<0.001). Post-hoc tests revealed the effect was not present for [^{18}F]flutemetamol cortical changes. Furthermore, CBF increases resulted in a higher $SUVr_{FLOW}$, while CBF decreases caused the opposite effect, except for [^{18}F]flutemetamol cortical changes (Figure 6). Varying the duration of the uptake period from 20 to 15 or 10 minutes provided essentially identical results for [^{18}F]flutemetamol. With respect to [^{18}F]florbetaben, $SUVr_{FLOW}$ calculated over 20 minutes showed significantly less CBF sensitivity compared with $SUVr_{FLOW}$ calculated over 10 minutes (p=0.005).

With respect to DVR_{FLOW_SRTM}, there was a small effect of linearization start time with later start times (corresponding to shorter acquisitions) being less sensitive to CBF changes ([^{18}F]flutemetamol: χ^2= 78, p<0.001, [^{18}F]florbetaben: χ^2= 8, p=0.04). More specifically, global and cortical CBF changes resulted in a small decrease in CBF sensitivity for later linearization times ([^{18}F]flutemetamol: χ^2= 66, p<0.001 and χ^2= 49, p<0.001, respectively; [^{18}F]florbetaben: χ^2=36 and p<0.001, χ^2=14 p=0.003, respectively). On the other hand, for [^{18}F]florbetaben, cerebellar CBF changes resulted in a small increase in the models sensitivity (χ^2=44, p<0.001; Supplementary Fig. 3).

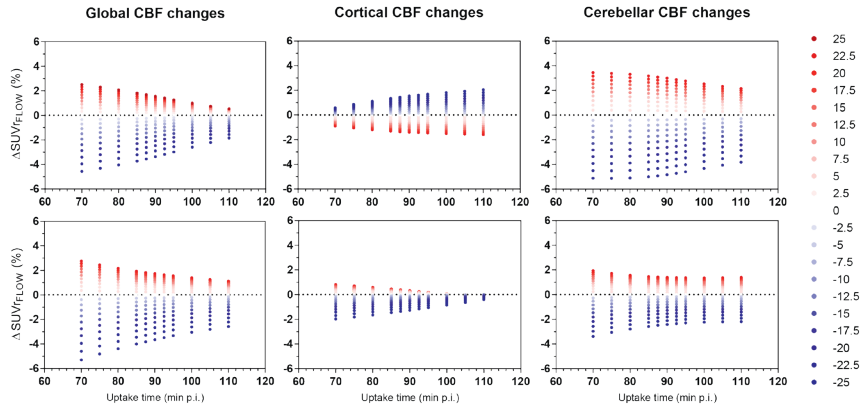


Figure 6. The effect of uptake time on sensitivity of SUVr to CBF changes. Upper row is [^{18}F]flutemetamol and bottom row is [^{18}F]florbetaben, depicted for intermediate amyloid (DVR=1.400 and DVR=1.538 for both tracers respectively) Red dots resemble CBF increases and blue dots CBF decreases.

Discussion

The current simulation study assessed the effect of global and regional CBF changes on regional amyloid quantification based on static, dynamic and dual-time window scanning-protocols using reference tissue approaches for [^{18}F]flutemetamol and [^{18}F]florbetaben. The results of the present simulation study indicate that, compared with DVR_{FLOW_SRTM} and $SUVr_{FLOW}$, DVR_{FLOW_RLOGAN} was least affected by changes in CBF for both [^{18}F]flutemetamol and [^{18}F]florbetaben. Furthermore, this sensitivity to CBF changes increased with increasing levels of amyloid. Finally, changes in the relative delivery rate R_1, obtained with SRTM, were highly correlated with simulated CBF changes.

The finding that DVR_{FLOW_RLOGAN} was robust against changes in CBF is in line with a previous study that found the same result for [^{11}C]PiB (*5*) and this robustness may be due to the linearity of the model (*24*). In contrast, SRTM was found to be somewhat more sensitive to changes in CBF for both tracers, which is in agreement with work of Bullich et al. (*7*). Their findings differ from the results reported here, in that they reported a slightly higher sensitivity of SRTM to cortical and a lower sensitivity to global CBF changes. This discrepancy may be attributed to differences in simulation design, the different cortical composite region and differences in fitter settings (such as parameter

boundary values) between the studies (*7,22*). In the present study, parameter boundary values were optimised to prevent fit parameters from becoming undetermined (i.e. k_2 and BP_{ND}), in particular for lower levels of amyloid.

The finding that DVR_{FLOW_SRTM} is sensitive to changes in CBF is probably due to the fact that the assumptions underlying SRTM are violated for both tracers, i.e. the requirement that tracer kinetics are well described by a one-tissue compartment model in both the target and reference tissue (*25*). More specifically, for both tracers the target and reference tissue kinetics have been shown to be better described by a two-tissue compartment model (*3,21*). In contrast, RLogan does not assume any specific number of target tissue compartments, robustly estimating DVR_{FLOW} independent of the underlying compartmental separation. Furthermore, most outliers were present for interval compared to full TACs (outliers interval TACs: [^{18}F]flutemetamol: 91.67%, [^{18}F]florbetaben: 66.67%). These outliers were only observed for DVR_{FLOW_SRTM} and in particular for TACs corresponding to low amyloid levels.

With respect to $SUVr_{FLOW}$, effects of CBF changes were relatively small for both tracers (see Figures 3 and 4, maximum 5.1 and 6.1% for [^{18}F]flutemetamol and [^{18}F]florbetaben, respectively). However, SUVr's sensitivity to global CBF changes increased for higher amyloid levels, as also seen for the other models. A comparable finding has been reported for [^{18}F]florbetapir, where amnestic MCI and AD subjects showed higher CBF sensitivity compared to controls, for large CBF decreases (-40%) (*18*). The maximum change in $SUVr_{FLOW}$ was within the maximum reported change for other amyloid tracers that assessed slightly more extreme flow variations (*5,18*). In addition, the sensitivity of $SUVr_{FLOW}$ to CBF changes was, as expected, dependent on acquisition start time, with later start times being more robust against CBF fluctuations, except for [^{18}F]flutemetamol cortical changes. For [^{18}F]flutemetamol, the direction of the relation between sensitivity to CBF changes and acquisition start time was, although less pronounced, comparable with [^{11}C]PiB findings. More specifically, for cortical CBF changes, an increased sensitivity was observed for later uptake-times, while this relationship was inverted for global changes (*5*). These results suggest that imaging data acquired at expedited instead of delayed uptake times should be interpreted carefully, as a greater impact of CBF changes on SUVr is expected.

It is of interest to note that, in some cases, the confounding effects of perfusion changes were different between the two tracers. TACs were simulated using parameters derived

from existing studies (*20,21*). Although sample sizes of those studies were different, it is unlikely that this had an effect on the final results, as both datasets consisted of 50% AD patients and 50% healthy controls. The most likely explanation for these differences in results is the difference in tracer kinetics of the two tracers, which can be deduced from the rate constants used for the simulations (Table 1). As mentioned above, reference tissue kinetics of both tracers were best described by including a second compartment, thereby violating the assumptions of a reference tissue model to some degree. More specifically, individual k_3 and k_4 parameters, as well as their ratio, differed between both tracers. This, in combination with differences in target tissue kinetics, could explain the observed differences between both tracers, which were most pronounced for SRTM, a model that assumes single tissue kinetics in both target and reference tissues. In other words, the difference in sensitivity to perfusion changes between both tracers is most likely due some degree of violation of underlying model assumptions, which may be different for the two tracers.

For both tracers, strong relationships ($R^2 \geq 0.95$) were observed between SRTM-derived R_1 changes and simulated changes in CBF. This finding was in agreement with the high correlation observed between SRTM-derived R_1 and $^{15}O\text{-}H_2O$ PET, reported by a combined $^{15}O\text{-}H_2O$ PET and [^{18}F]florbetapir study (*16*). More specifically, it suggests that for both [^{18}F]flutemetamol and [^{18}F]florbetaben, R_1 could be used as a proxy for measuring CBF. Finally, the effect of noise (resembling regions of interest) on sensitivity to CBF changes for DVR_{FLOW_RLOGAN}, DVR_{FLOW_SRTM} and $SUVr_{FLOW}$ was limited across amyloid levels. The effect of noise corresponding to a voxel level needs further validation, preferably using parametric imaging data.

As mentioned before, the main goal of this work was to assess the effect of global and regional CBF changes on regional amyloid quantification for [^{18}F]flutemetamol and [^{18}F]florbetaben studies. Overall, a maximum CBF induced change in amyloid outcome measure of 13.4% (SRTM) was observed across methods. However, it must be noted that the evaluated range of CBF changes was more extreme than the change one would typically expect during the timespan of a clinical trial or longitudinal study from an ageing or early AD perspective. On the other hand, in late AD stages the CBF changes may be larger and drugs may also exert effects on CBF (*12,13,15*). Although the exact size of the effect is unknown for many drugs, the effect of regularly used substances such as caffeine and nicotine ranges from -22 to +25% (*28,29*). Therefore, besides the effect of drugs or stimulants, effects of CBF changes on DVR_{FLOW} and $SUVr_{FLOW}$ will likely be

smaller than the effects reported in the present study. The present results suggest that if large changes in amyloid load are to be expected, the possible confounding effects due to changes in perfusion are likely to be insignificant, obviating the need for a dynamic scanning protocol. However, in drug development studies, the drug may have unknown effects on perfusion, which must be well understood prior to implementation of simplified methods based on a static scanning protocol. This is especially relevant in secondary prevention trials, where measured changes in amyloid load may be small. In those cases a dynamic or dual-time window protocol should be used to assess whether changes in perfusion have occurred and to estimate their magnitude. Based upon this information, it can be decided whether a static protocol is sufficient or a dynamic protocol is needed to address a specific clinical or research question. Finally, it should be noted that a dynamic or dual-time window scanning protocol also can provide a measurement of relative perfusion, which may be an additional relevant source of information.

It should be noted that an exhaustive assessment of additional sources of bias was outside the scope of the present study. However, it is known from previous studies that SUVr typically shows a positive bias, while RLogan tends to underestimate the gold standard, and these intrinsic biases challenge the pooling of data from different methods (*5,7,20,21*). Moreover, other factors such as image contrast and test-retest reliability may play a role when deciding on the method of choice for analysing data (*6*). This implies that additional research, in particular test-retest and parametric imaging studies, is warranted.

Finally, given that the linear model (RLogan) outperformed the non-linear model (SRTM) with respect to robustness against CBF changes, one could conclude that SRTM is not the optimal model for describing the tracer's kinetics. Alternatively, a linearized form, basis function implementation of SRTM (RPM) (*30*) could be evaluated, which has shown improved performance, especially for noisy data, compared to the original implementation of SRTM for [^{18}F]flutemetamol (*20*). The present study focussed on data corresponding to a region of interest level. As a next step, future studies could validate the application of parametric methods, such as RPM and SRTM2 (*31*), against the present results. The implementation of these methods (for example RPM's basis function settings) would ideally be validated using imaging data.

Conclusion

RLogan was least affected by changes in cerebral blood flow and is therefore, at least within this context, the preferred method for regional amyloid quantification of [^{18}F]flutemetamol and [^{18}F]florbetaben. The same data could be analysed again using SRTM in order to determine the relative rate of delivery R_1, which showed good correlation with CBF. Finally, in most cases, effects of CBF changes on SUVr were relatively small, in particular for later uptake-times.

Acknowledgements

The authors would like to thank AMYPAD for all support and in particular prof. Frederik Barkhof for facilitating this research. In addition, we would like to thank Christopher Foley from GE for his feedback and the staff of the department of Nuclear Medicine, University of Leipzig for skilful acquisition and analyses of the [^{18}F]florbetaben data during the clinical development.

Note. Published online by SAGE publishers on the 11th of April 2020, doi: 10.1177/0271678X20918029

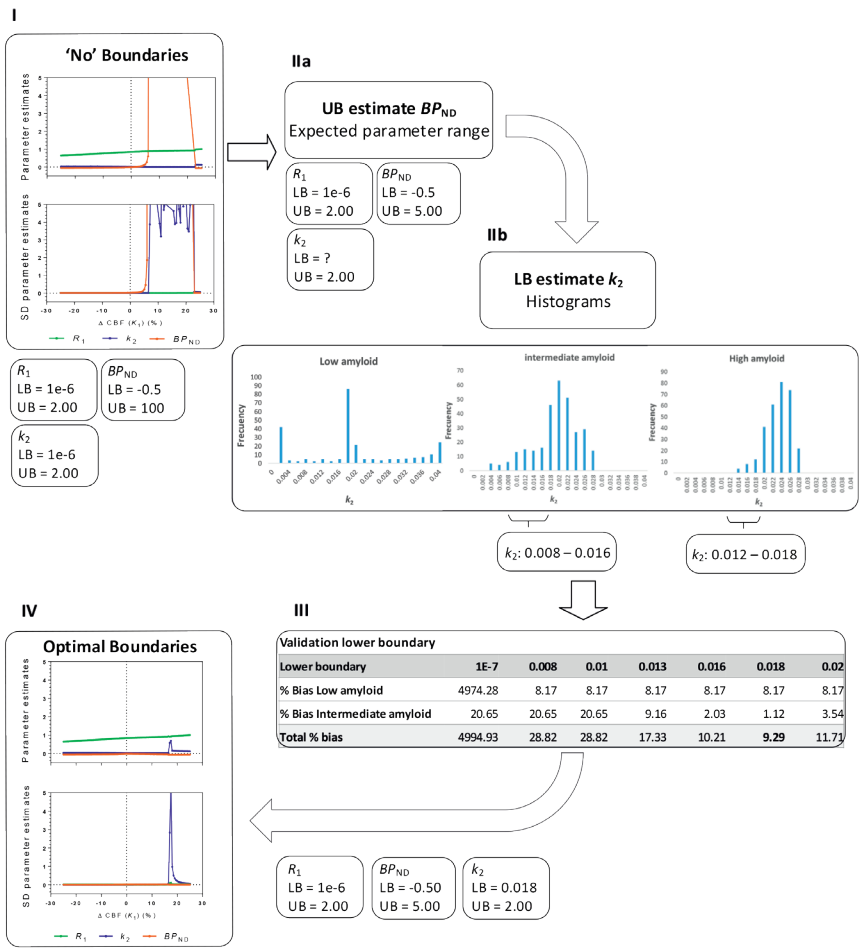
References

1. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010;9:119-128.
2. Mallik A, Drzezga A, Minoshima S. Clinical Amyloid Imaging. Semin Nucl Med. 2017;47:31-43.
3. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 Study of the Pittsburgh Compound B Derivative 18F-Flutemetamol in Healthy Volunteers and Patients with Probable Alzheimer Disease. Journal of Nuclear Medicine. 2009;50:1251-1259.
4. Carson RE, Channing MA, Blasberg RG, et al. Comparison of Bolus and Infusion Methods for Receptor Quantitation: Application to [18 F]Cyclofoxy and Positron Emission Tomography. Journal of Cerebral Blood Flow & Metabolism. 1993;13:24-42.
5. van Berckel BNM, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. Journal of Nuclear Medicine. 2013;54:1570-1576.
6. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008;42:76-86.
7. Bullich S, Barthel H, Koglin N, et al. Validation of Noninvasive Tracer Kinetic Analysis of 18F-Florbetaben PET Using a Dual-Time-Window Acquisition Protocol. J Nucl Med. 2018;59:1104-1110.
8. Heeman F, Yaqub M, Lopes Alves I, et al. Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. EJNMMI Research. 2019;9:32.
9. Frackowiak RS. Measurement and imaging of cerebral function in ageing and dementia. Prog Brain Res. 1986;70:69-85.
10. Pantano P, Baron JC, Lebrun-Grandié P, Duquesnoy N, Bousser MG, Comar D. Regional cerebral blood flow and oxygen consumption in human aging. Stroke. 1984;15:635-641.
11. Bremmer JP, van Berckel BNM, Persoon S, et al. Day-to-day test-retest variability of CBF, CMRO2, and OEF measurements using dynamic 15O PET studies. Mol Imaging Biol. 2011;13:759-768.
12. Mosca F, Bray M, Lattanzio M, Fumagalli M, Tosetto C. Comparative evaluation of the effects of indomethacin and ibuprofen on cerebral perfusion and oxygenation in preterm infants with patent ductus arteriosus. The Journal of Pediatrics. 1997;131:549-554.
13. de Jong DLK, de Heus RAA, Rijpma A, et al. Effects of Nilvadipine on Cerebral Blood Flow in Patients With Alzheimer Disease: A Randomized Trial. Hypertension. 2019;74:413-420.
14. Alegret M, Cuberas-Borrós G, Vinyes-Junqué G, et al. A Two-Year Follow-Up of Cognitive Deficits and Brain Perfusion in Mild Cognitive Impairment and Mild Alzheimer's Disease. Journal of Alzheimer's Disease. 2012;30:109-120.

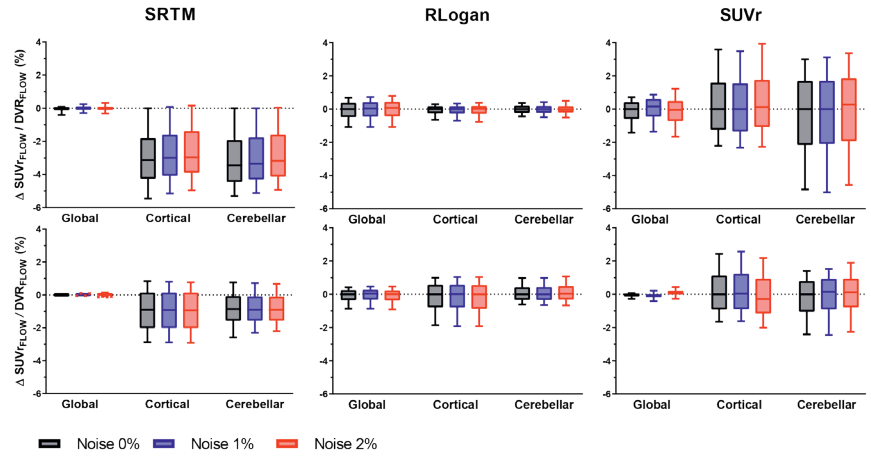
15. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. 18F-FDG PET, the early phases and the delivery rate of 18F-AV45 PET as proxies of cerebral blood flow in Alzheimer's disease: Validation against 15O-H2O PET. Alzheimer's & Dementia. August 2019.
16. Postiglione A, Lassen NA, Holman BL. Cerebral blood flow in patients with dementia of Alzheimer's type. Aging Clin Exp Res. 1993;5:19-26.
17. Sojkova J, Beason-Held L, Zhou Y, et al. Longitudinal Cerebral Blood Flow and Amyloid Deposition: An Emerging Pattern? Journal of Nuclear Medicine. 2008;49:1465-1471.
18. Ottoy J, Verhaeghe J, Niemantsverdriet E, Engelborghs S, Stroobants S, Staelens S. A simulation study on the impact of the blood flow-dependent component in [18F]AV45 SUVR in Alzheimer's disease. Gelovani JG, ed. PLOS ONE. 2017;12:e0189155.
19. Chen YJ, Rosario BL, Mowrey W, et al. Relative 11C-PiB Delivery as a Proxy of Relative CBF: Quantitative Evaluation Using Single-Session 15O-Water and 11C-PiB PET. J Nucl Med. 2015;56:1199-1205.
20. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18)F]flutemetamol. Neuroimage. 2015;121:184-192.
21. Becker GA, Ichise M, Barthel H, et al. PET quantification of 18F-florbetaben binding to β-amyloid deposits in human brains. J Nucl Med. 2013;54:723-731.
22. Yaqub M, Boellaard R, Kropholler MA, Lammertsma AA. Optimization algorithms and weighting factors for analysis of dynamic PET studies. Phys Med Biol. 2006;51:4217-4232.
23. Morris ED, Endres CJ, Schmidt KC, Christian BT, Muzic RF, Fisher RE. Kinetic Modeling in Positron Emission Tomography. Emission Tomography: The Fundamentals of PET and SPECT. November 2004:499-540.
24. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. J Cereb Blood Flow Metab. 1996;16:834-840.
25. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996;4:153-158.
26. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association. 1952;47:583-621.
27. Tukey JW. Exploratory data analysis. Reading, Mass.: Addison-Wesley Pub. Co.; 1977.
28. Lunt MJ, Ragab S, Birch AA, Schley D, Jenkinson DF. Comparison of caffeine-induced changes in cerebral blood flow and middle cerebral artery blood velocity shows that caffeine reduces middle cerebral artery diameter. Physiol Meas. 2004;25:467-474.
29. Wennmalm A. Effect of cigarette smoking on basal and carbon dioxide stimulated cerebral blood flow in man. Clin Physiol. 1982;2:529-535.

30. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. NeuroImage. 1997;6:279-287.
31. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. J Cereb Blood Flow Metab. 2002;22:1440-1452.

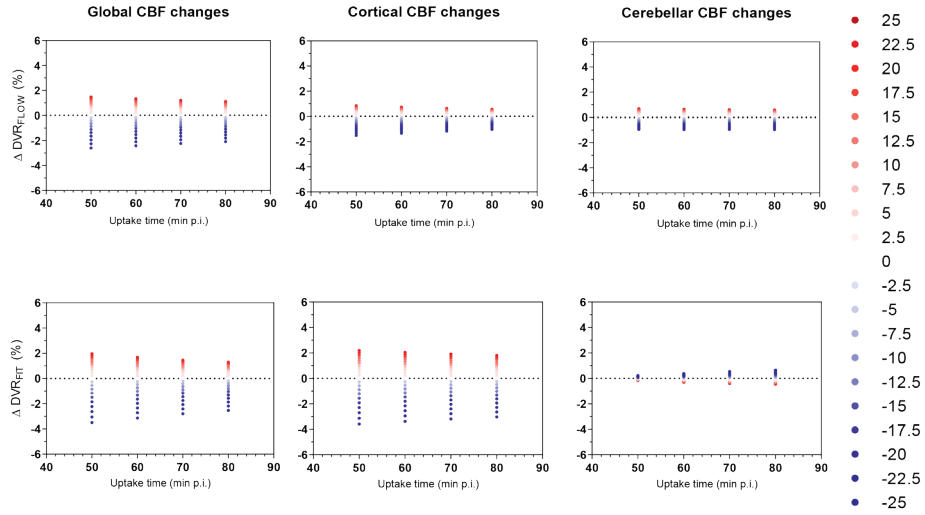
Supplementary materials



Supplementary Figure 1. Boundary validation I. Example of SRTM derived parameters for noiseless full TACs with cortical CBF changes, corresponding to low amyloid. The fitting routine required implementation of boundaries (zero or infinite values not possible), therefore the following start-parameters were used: R_1(1e-6; 2), k_2(1e-6; 2), BP_{ND}(-0.5; 100). II. BP_{ND} and k_2 were poorly determined; BP_{ND} values were enormous while k_2 values approached zero. (**a**) BP_{ND} upper boundary estimates were adjusted based upon the simulated parameter range and (**b**) k_2 lower boundaries were adjusted based upon parameter histograms across binding levels. III. Lower boundary values were fine-tuned until the lowest overall bias (calculated from the simulated BP_{ND}+1) was obtained. IV. Same data as in I, but with new boundaries.
Note: All example values/graphs correspond to the [18F]flutemetamol tracer, for [18F]florbetaben the exact same procedure was carried out.



Supplementary Figure 2. The effect of noise on sensitivity to CBF changes across methods. Upper row is [18F]flutemetamol and bottom row is [18F]florbetaben. Note: plots correspond to the lowest level of amyloid load (DVR=1.022, DVR=1.026, for both tracers, respectively. Whiskers were defined according to the Tukey method and outliers are depicted as red crosses.



Supplementary Figure 3. The effect of linearization start-time on sensitivity to CBF changes for RLogan. Upper row is [18F]flutemetamol and bottom row is [18F]florbetaben. Results correspond to an intermediate level of amyloid (DVR=1.400 and DVR=1.538 for both tracers respectively) Red dots resemble CBF increases and blue dots CBF decreases.

7.

Impact of cerebral blood flow and amyloid load on SUVR bias

Fiona Heeman, Maqsood Yaqub, Janine Hendriks, Bart N.M. van Berckel, Lyduine E. Collij, Katherine R. Gray, Richard Manber, Robin Wolz, Valentina Garibotto, Catriona Wimberley, Craig Ritchie, Frederik Barkhof, Juan Domingo Gispert, David Vállez García, Isadora Lopes Alves, Adriaan A. Lammertsma
On behalf of the AMYPAD Consortium

Submitted

Abstract

Despite the widespread use of static amyloid PET scans for clinical and research purposes, it is known that the standardized uptake value ratio (SUVR), calculated from a static scan, may be biased compared with the distribution volume ratio (DVR) derived from a dynamic scan. This bias may be partially explained by changes in cerebral blood flow (CBF) and is likely to be also dependent on the severity of the underlying amyloid-β (Aβ) pathology. To date, most Alzheimer's disease (AD) studies have compared SUVR and DVR only at a diagnostic group level and not as a function of underlying Aβ pathology. The purpose of the present study was to compare SUVR with DVR and to evaluate the effects of underlying Aβ pathology and CBF on bias in SUVR in a group of mainly cognitively unimpaired participants (*N*=121).

Methods: Participants were scanned according to a dual-time window protocol, with either [^{18}F]flutemetamol (*N*=90) or [^{18}F]florbetaben (*N*=31). The validated voxel-based implementation of the two-step simplified reference tissue model was used to derive DVR and R_1 and SUVR was calculated for a 90-110 min post-injection uptake window, all with the cerebellar grey matter as reference tissue. First, linear regression and Bland-Altman analyses were used to compare SUVR with DVR. Then, Generalized Linear Models were applied to evaluate whether SUVR or bias in SUVR relative to DVR ($SUVR_{bias}$) could be explained by R_1 for four regions (i.e. global cortical average (GCA), precuneus, posterior cingulate, and orbitofrontal cortex).

Results: Despite high correlations (GCA: $R^2 \geq 0.85$), large overestimation and proportional bias of SUVR relative to DVR was observed (regression line of the proportional bias: GCA: [^{18}F]flutemetamol R^2=0.69, slope=0.56 and intercept=-0.48; [^{18}F]florbetaben: R^2=0.65, slope=0.31 and intercept=-0.20). Negative associations were observed between both SUVR or $SUVR_{bias}$ and $R_{1,}$ albeit non-significant.

Conclusion: The present findings demonstrate that bias in SUVR relative to DVR is primarily due to underlying Aβ pathology. Furthermore, in a cohort consisting mainly of cognitively unimpaired individuals, the effect of CBF on bias in SUVR appears limited.

Key words: Alzheimer's disease, amyloid PET, cerebral blood flow, quantification, SUVR bias

Introduction

At present, amyloid positron emission tomography (PET) imaging is routinely used in specialised memory clinics to support the diagnosis of Alzheimer's disease (AD) dementia. Routinely, a static acquisition is performed and the resulting image is used to visually determine the presence or absence of amyloid-beta (Aβ) plaques by a trained nuclear medicine physician. In addition to their clinical use, static scans have also been widely used for research purposes, mainly because of their technical and logistical simplicity. From these scans, the commonly used standardized uptake value ratio (SUVR) can be calculated, by simply dividing tracer uptake in the target by that in a reference region. Nevertheless, previous studies have demonstrated that SUVR can be biased compared with the non-displaceable binding potential (BP_{ND}) or distribution volume ratio (DVR = BP_{ND}+1) derived from a dynamic scan (*1*–*3*). In particular, SUVR can be affected by changes in cerebral blood flow (CBF), tracer clearance, and/or extraction fraction (*1*,*2*,*4*). These effects have been illustrated in a longitudinal observational study that reported a decrease in [^{11}C]PiB SUVR, while DVR remained unchanged (*2*), and accompanying simulations demonstrated that this decrease in SUVR was likely related to a decrease in CBF.

The degree of bias in SUVR also depends on the underlying level of pathology (*1*) and, in the case of AD, it is known that disease stage and underlying Aβ pathology are not linearly correlated (*5*). Yet, most AD studies have compared SUVR and DVR only at a diagnostic group level and, therefore, the relationship between bias and extent of underlying Aβ pathology is not fully characterised (*6*–*8*).

Over the last decade, the focus of amyloid PET studies has shifted to the early phases of AD, emphasizing the importance of understanding the bias in SUVR in this pre-clinical group (*9*–*12*). In particular, considering that confounding factors such as changes in CBF, tracer clearance, and extraction fraction may not be constant throughout the disease, their impact on SUVR could also differ in this pre-clinical population compared with AD dementia patients (*1*,*13*,*14*). Therefore, the purpose of the present study was to investigate bias in SUVR relative to DVR in elderly individuals and assess to what extent this bias can be explained by factors such as underlying Aβ pathology and CBF (as measured by relative tracer delivery) for both [^{18}F]flutemetamol and [^{18}F]florbetaben.

Materials and Methods

Participants

The study included *N*=121 participants from the AMYloid imaging to Prevent Alzheimer's Disease (AMYPAD) Prognostic and Natural History Study (PNHS) (*15*), who were scanned across four different centres: The University of Edinburgh (UEDIN), Barcelona Beta Brain Research Center (BBRC), Amsterdam University Medical Centres, location VUmc (Amsterdam UMC, VUmc) and University of Geneva (UNIGE). All participants had at least one PET and one T1-weighted MR scan available, and visual assessment of the PET scans was performed locally by a trained nuclear medicine physician according to the manufacturer's reading guidelines (*16*). In addition, participants underwent standard neurological screening and neuropsychological assessment (e.g. mini-mental state examination, MMSE) and information regarding their *APOE*-ε4 status was available. Before participating in the study, all participants provided written informed consent in accordance with the Declaration of Helsinki. Study protocols were approved by all local Medical Ethics Review Committees.

Image acquisition

Prior to each PET scan, an MR sequence (MRAC) or low dose CT (ldCT) was acquired for attenuation correction purposes (supplementary Table 1 specifies which centres used an MR and which a CT). Participants then received a bolus injection of either [^{18}F]flutemetamol (*N*=90, 186±10 MBq) or [^{18}F]florbetaben (*N*=31, 283±20 MBq) and underwent a dynamic PET scan according to a dual-time window protocol (*17*). This scanning protocol consisted of an early dynamic scan from 0 to 30 minutes post-injection (p.i.) followed by a break of 60 minutes, and then a late dynamic scan from 90 to 110 minutes p.i. For each scanner, recommended clinical reconstruction settings for that scanner were used and scans were reconstructed into 22 frames (6x5, 3x10, 4x60, 2x150, 2x300, and 1x600 seconds) for the early scan, and four frames (4x300 seconds) for the late scan. All sites applied the following corrections: detector normalisation, dead time, attenuation, scatter, decay and randoms. Site-specific reconstruction methods can be found in supplementary Table 1.

Image QC and processing

First, a quality control (QC) check of the image meta-data was performed (i.e. frame start times, duration, and the number of frames) and the presence of motion between PET scans, and corresponding MRAC or ldCT, was evaluated. In the case of severe motion

(>5 mm), these scans were excluded from analyses. In addition, it was verified whether the whole brain was in the field of view and whether no technical errors occurred, such as a delay in scanner start time. Next, presence of between-frame motion was assessed visually. Then, both early and late phases of the PET scan were coregistered to the T1-weighted MR scan using rigid registration and the resulting PET scans were resampled to have the same resolution as the T1-weighted MR. This was done to prevent changes to the region of interest (ROI) template images, which were already in the same space as the T1-weighted MR scan. Following this step, coregistered PET images were combined and the resulting scan was divided into five blocks of 10 minutes duration (A: frames 1-15, B: frames 16-17, C: frame 18, D: frames 19-20, and E: frames 21-22). Presence of motion between blocks was assessed visually, by comparing whether contours between blocks were overlapping. Next, using rigid co-registration (based on Elastix software), blocks were co-registered and the maximum translation in each direction (x-y-z) was determined (*18*,*19*). Based on a comparison between visual assessment and required translation, it was determined whether motion exceeded 1 mm translation and should be corrected for (using Elastix-based rigid co-registration). Subsequently, subject-specific regions of interest (ROIs) generated using Learning Embeddings for Atlas Propagation (LEAP) (*20*) (in T1-weighted MR space) were applied to the combined PET scan to extract a reference tissue time-activity curve (TAC) of the cerebellar grey matter.

Parametric analysis

Interpolation of the missing data points of the reference tissue TAC (corresponding to the 60 min break) was performed using the reversible two tissue compartment model (4 rate constants) with additional blood volume fraction parameter (2T4k_V_b) and a tracer-specific plasma input function as described previously (*17*). Visual QC of the interpolated reference tissue TAC was performed to assess whether the interpolation of the missing data points had a smooth connection with the measured data. In the case of sub-optimal interpolation (i.e. clear discontinuity between interpolated and measured data), cubic interpolation was used as alternative (*N*=8). Next, parametric modelling was performed using the voxel-based implementation of the two-step simplified reference tissue model (SRTM2) (*21*), as implemented in the PPET software package (*22*) and validated previously (*23*), to compute parametric BP_{ND} and relative tracer delivery (R_1) images. For SRTM2, k_2' was determined by taking the median k_2' across all voxels with a BP_{ND} higher than 0.05 from a first run using receptor parametric mapping (RPM) (*24*,*25*). In addition, SUVR was calculated from 90-110 minutes p.i. To allow for comparability between metrics, DVR was calculated as BP_{ND}+1. Finally, the subject-specific global

cortical average (GCA) template (from the Centiloid (CL) method), which is a composite region consisting of frontal, temporal, parietal, and insular cortices, precuneus, and striatum (*26*), was applied to the parametric data to obtain global SUVR and DVR values. In addition, the subject-specific ROIs generated using LEAP were used to extract values for regions that are known to show early accumulation of Aβ plaques: precuneus, posterior cingulate cortex (PCC), and orbitofrontal cortex (OFC) (*12*,*27*,*28*). For all participants, Centiloid values were available as computed via the standard CL processing pipeline, which allows for between-tracer comparisons (*26*,*29*)

Statistical analysis

All statistical analyses were performed in R (version 4.0.3; R Foundation for Statistical Computing, Vienna, Austria) and stratified per tracer. A statistically significant result was defined as $p< 0.05$, no corrections for multiple comparisons were applied.
First, demographic differences between Aβ-positive and negative participants (based on a visual assessment) were investigated using *t*-tests (or in the case of non-normal distribution, Mann-Whitney *U* tests) and chi-square tests.

Relationship between SUVR and DVR
Linear regression analyses were used to assess the relationships between SUVR and DVR. In addition, Bland-Altman analyses (*30*) were used to assess potential bias between these metrics, and the presence of proportional bias was determined visually for the GCA and three early regions. When present, proportional bias was further evaluated by fitting a regression line through the Bland-Altman plot.

Relationship between (bias in) SUVR and relative CBF
First, bias in SUVR relative to DVR ($SUVR_{bias}$) was calculated ((SUVR-DVR)/DVR*100%). Then, Generalised Linear Models (GLMs) were constructed to understand whether the relative tracer delivery (R_1) could explain the remaining variance of either SUVR or $SUVR_{bias}$ beyond the main predictor, DVR. Note, the covariate "centre" is only included for [^{18}F]flutemetamol.

$$SUVR \sim DVR + R_1 + DVR * R_1 + Age + Sex + APOE\text{-}\varepsilon4 + centre$$
$$SUVR_{bias} \sim DVR + R_1 + DVR * R_1 + Age + Sex + APOE\text{-}\varepsilon4 + centre$$

Before applying the GLMs to the data, all predictors (DVR, R_1, and DVR*R_1) and covariates (age, sex, *APOE*-ε4 carriership, and centre in the case of [^{18}F]flutemetamol)

were correlated with each other to check for collinearity. This was done using Pearson's correlation, point-biserial correlation, and Goodman Kruskal' lambda for continuous, continuous, and categorical or categorical variables, respectively. In the case of high overlap ($r\geq0.70$) (*31*), the variables were correlated with the dependent variable (SUVR and $SUVR_{bias}$) to determine which one should be deleted from the model to remove redundancy, i.e. the one with the lowest correlation. All analyses with the final model were performed using the GCA, as well as with the three early regions.

Results

Participants

Participant characteristics are shown in Table 1. Most participants were cognitively unimpaired (*N*=104), while a minority (*N*=17) showed impairment to some extent, based upon the clinical dementia rating (CDR) score (CDR≥0.5). For participants scanned with [^{18}F]flutemetamol, 21.1% (19/90) were Aβ-positive (based on visual assessment of the scan), while for participants scanned with [^{18}F]florbetaben, this was 38.7% (12/31). Furthermore, Aβ-positive participants scanned with [^{18}F]flutemetamol were slightly older than the Aβ-negative group. As expected, the proportion of *APOE*-ε4 carriers and the overall amyloid burden (in CL, SUVR and DVR units) were higher among Aβ-positive compared with Aβ-negative participants for both tracers, while no differences were observed in R_1 (Figure 1).

Table 1. Participant demographics

	[^{18}F]flutemetamol			[^{18}F]florbetaben		
	All (*N*=90)	**Aβ-negative** (*N*=71)	**Aβ-positive** (*N*=19)	**All** (*N*=31)	**Aβ-negative** (*N*=19)	**Aβ-positive** (*N*=12)
Age	67.2±6.3	66.2±6.3**	70.7±4.7	67.3±7.4	66.0± 6.7	69.5±8.2
Females (%)	54.4	57.7	42.1	67.7	73.7	58.3
MMSE	28.9±1.4	29.2±1.1	28.2±2.1	29.1±1.3	29.4±0.8	28.7±1.8
***APOE*-ε4+ (%)**	41.6	34.3*	68.4	35.5	15.8*	66.7
Centiloid	17.85±25.8	8.20± 12.0**	53.70±32.2	32.24±28.8	16.60±14.3**	57.01±36.54
SUVR	1.51±0.3	1.41±0.1**	1.86±0.3	1.48±0.2	1.36±0.1**	1.66±0.3
DVR	1.23±0.1	1.18±0.1**	1.41±0.2	1.26±0.2	1.18±0.1**	1.40±0.2
R_1	1.00±0.1	1.00±0.1	0.97±0.1	0.96±0.2	0.96±0.1	0.95±0.1
Centre	A,B,C,D	A,B,C,D	A,B,C	A	A	A
Inj. dose (MBq)	186.2±10.4	188.8±10.1	184.0±7.7	282.9±19.7	284.4±17.9	280.5±22.9

Values depicted as mean ± SD, MMSE=mini-mental state examination, Inj. dose = Net injected dose *p<0.05, **p<0.01, compared with the Aβ-positive group. A=Amsterdam UMC, B=BBRC, C=UNIGE, D=Edinburgh

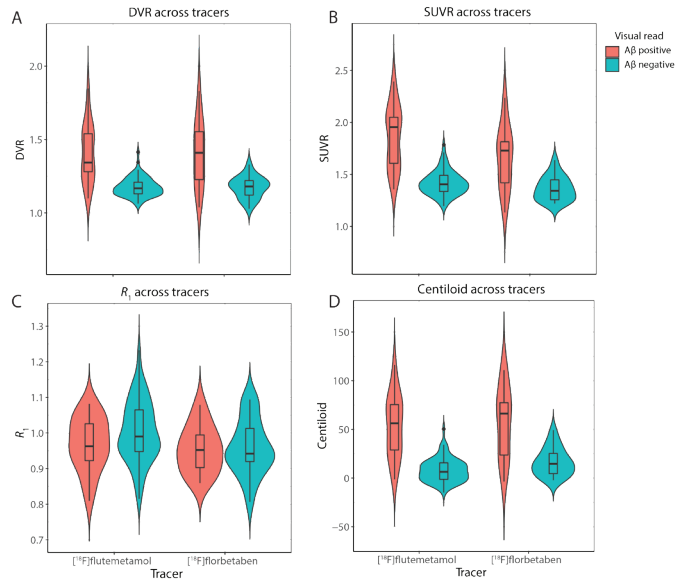


Figure 1. Parameter distribution across tracers
Violin plots showing the distribution of R_1, DVR, SUVR and Centiloid for Aβ-positive and negative scans. Small boxplots inside the violin plots display median and quartile range of the distribution.

Relationship between SUVR and DVR

High correlations were observed between global SUVR and DVR ([^{18}F]flutemetamol: R^2=0.85 [^{18}F]florbetaben: R^2=0.95, Figure 2). The same was true for the early regions, with small differences between regions ([^{18}F]flutemetamol precuneus: R^2=0.85, OFC: R^2=0.86, PCC: R^2=0.82, [^{18}F]florbetaben precuneus: R^2=0.93, OFC: R^2=0.87, PCC: R^2=0.85).

SUVR showed an overestimation compared with global DVR for both tracers ([^{18}F]flutemetamol mean bias: 0.29, 95% Limits of Agreement (LoA) 0.03 to 0.54, and [^{18}F]florbetaben mean bias: 0.22 95% LoA 0.05 to 0.38), which was strongly proportional to the underlying amyloid burden, being slightly larger for [^{18}F]flutemetamol ([^{18}F]flutemetamol: R^2=0.70, slope=0.56 and intercept=-0.48, [^{18}F]florbetaben: R^2=0.65, slope=0.31 and intercept=-0.20, Figure 2). For each of the early regions, a similar pattern in terms of overestimation and proportional bias was observed, although for most early regions, the proportionality was slightly stronger ([^{18}F]flutemetamol precuneus: R^2=0.81, slope=0.74 and intercept=-0.79, OFC: R^2=0.74, slope=0.59 and intercept=-0.54, PCC: R^2=0.73, slope=0.66 and intercept=-0.56, and [^{18}F]florbetaben precuneus: R^2=0.70, slope=0.39 and intercept=-0.32, OFC: R^2= 0.44, slope=0.31 and intercept=-0.23, PCC: R^2=0.61, slope=0.48 and intercept=-0.35).

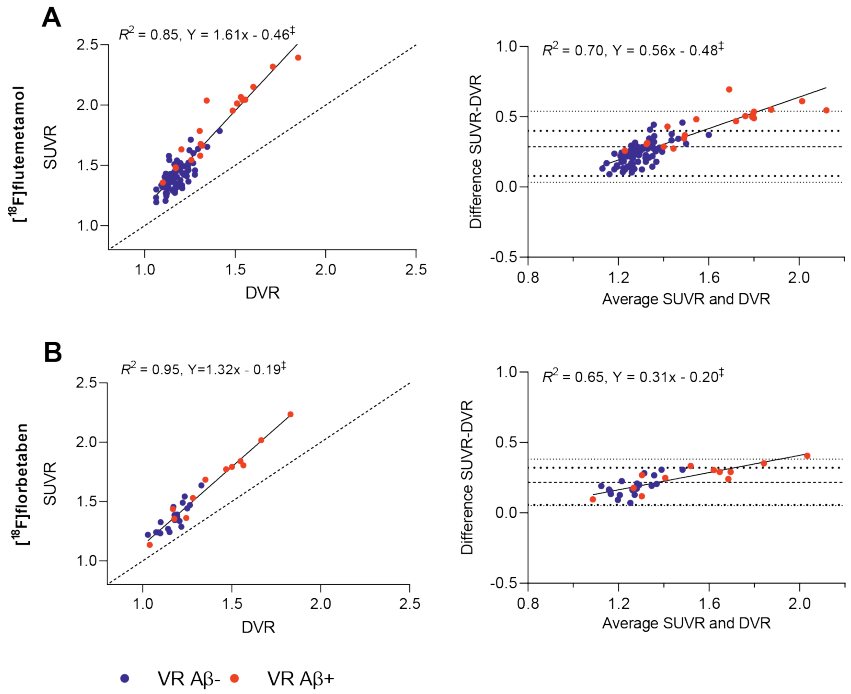


Figure 2. Relationship between DVR and SUVR
Correlation and Bland-Altman plots to assess the relationship between global cortical DVR and SUVR for **(a)** [18F]flutemetamol and **(b)** [18F]florbetaben. Dotted lines corresponds to 95% Limits of Agreement. ***$p < 0.001$, VR: visual read.

Relationship between independent variables

The univariate correlation analysis between all predictor (DVR, R_1, and DVR*R_1) and covariate variables (age, sex, *APOE*-ε4 carriership, and centre in the case of [18F]flutemetamol) demonstrated high correlations only between DVR and DVR*R_1 (range *r*: 0.70-0.95). Comparisons with the dependent variables showed higher correlations between DVR and SUVR or $SUVR_{bias}$ (range *r*: 0.91-0.97, 0.24-0.60, respectively) than between DVR*R_1 and SUVR or $SUVR_{bias}$ (range *r*: 0.55-0.90, 0.13-0.46, respectively). Therefore, GLMs were constructed without interaction terms, which resulted in the following models (1a and b). Note, the covariate "centre" is only included for [18F]flutemetamol.

$$SUVR \sim DVR + R_1 + Age + Sex + APOE\text{-}\varepsilon4 + centre \quad (1a)$$

$$SUVR_{bias} \sim DVR + R_1 + Age + Sex + APOE\text{-}\varepsilon4 + centre \quad (1b)$$

Relationship between SUVR and relative CBF

As expected, there was a significant relationship between SUVR and DVR for both tracers across regions (range coefficient estimate [18F]flutemetamol: 1.60-1.84, [18F]florbetaben: 1.31-1.43, $p<0.001$, Table 2a and b, supplementary Table 2a-f). In addition, for both tracers there were negative associations between SUVR and R_1 for the GCA (coefficient estimate [18F]flutemetamol: -0.20, [18F]florbetaben: -0.17, Figure 3), PCC (coefficient estimate [18F]flutemetamol: -0.20, [18F]florbetaben: -0.21) and, in the case of [18F]flutemetamol, for the precuneus (coefficient estimate: -0.20), albeit not statistically significant (*ns*) (Table 2a and b, supplementary Table 2a,b,d). For [18F]flutemetamol specifically, a statistically significant association was also observed between SUVR and centre UNIGE, for the GCA and OFC (coefficient estimate: -0.13 and 0.16, for both regions respectively, $p<0.05$, Table 2a, supplementary Table 2c). No statistically significant associations were observed between SUVR and the other independent variables.

Table 2a. Relationship between (bias in) SUVR and independent variables for [18F]flutemetamol

GCA	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.70#	1.55 - 1.85	32.05#	19.78 – 44.31
R_1	-0.20	-0.43 - 0.04	-14.63	-34.12 – 4.86
Age	0.00	0.00 - 0.01	0.17	-0.10 – 0.44
Sex	0.00	-0.04 - 0.03	-0.18	-3.17 – 2.81
***APOE*-ε4**	0.03	-0.01 - 0.07	2.72	-0.30 – 5.75
BBRC	0.00	-0.07 - 0.08	0.31	-6.26 – 6.88
Amsterdam UMC	-0.08	-0.16 - 0.00	-6.32	-13.21 – 0.58
UNIGE	-0.13*	-0.24 --0.02	-10.16*	19.18 –-1.14

Females, *APOE*-ε4 non-carriers and centre UEDIN were used as reference groups. GCA: global cortical average, *$p<0.05$, **$p<0.01$, #$p<0.001$

Table 2b. Relationship between (bias in) SUVR and independent variables for [18F]florbetaben

GCA	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.31#	1.19 - 1.44	9.46	-0.78 - 19.69
R_1	-0.17	-0.49 - 0.14	-13.43	-39.00 - 12.14
Age	0.00	0.00 - 0.00	0.10	-0.16 - 0.37
Sex	0.00	-0.05 - 0.05	-0.12	-4.19 - 3.95
***APOE*-ε4**	0.02	-0.03 - 0.07	2.00	-1.96 - 5.97

Females and *APOE*-ε4 non-carriers were used as reference groups. GCA: global cortical average, *$p<0.05$, **$p<0.01$, #$p<0.001$

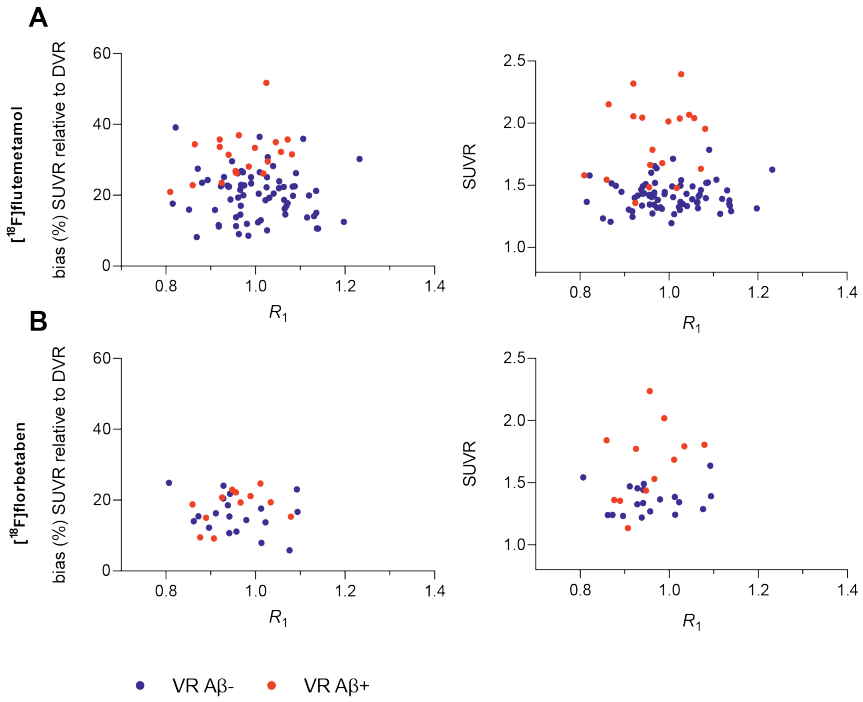


Figure 3. Relationship between global cortical relative tracer delivery (R_1) and amyloid burden
Correlation plots shown for **(a)** [18F]flutemetamol and **(b)** [18F]florbetaben. VR: visual read.

Relationship between bias in SUVR and relative CBF

For both tracers, positive associations were reported between $SUVR_{bias}$ and DVR across regions (range coefficient estimate [18F]flutemetamol: 13.54-33.49, $p<0.05$ and [18F]florbetaben: 9.46-17.94, statistically significant only for the precuneus $p<0.05$, Table 2a and b, supplementary Table 2a-f). In addition, negative associations were observed between $SUVR_{bias}$ and R_1 for the GCA (coefficient estimate [18F]flutemetamol: -14.63, [18F]florbetaben: -13.43, Figure 3), PCC (coefficient estimate [18F]flutemetamol: -9.99, [18F]florbetaben: -17.63) and, in the case of [18F]flutemetamol, also for the precuneus and OFC (coefficient estimate: -8.96 and -6.77, for both regions respectively), albeit not statistically significant (Table 2a and b, supplementary Table 2a-c,e). For [18F]flutemetamol, significant, positive associations were observed between $SUVR_{bias}$ and *APOE*-ε4 (coefficient estimate precuneus: 3.30, $p<0.05$) and age (coefficient estimate PCC: 0.24, $p<0.05$), although with small coefficient estimates. Finally, associations between $SUVR_{bias}$ and centre UNIGE and Amsterdam UMC were observed for the

GCA (coefficient estimate: 6.32, *ns* and -10.16, $p<0.05$) and OFC (coefficient estimate: 5.99 and -12.17, *ns*), for both centres respectively (Table 2a, supplementary Table 2c). No statistically significant associations were observed between $SUVR_{bias}$ and the other independent variables.

Discussion

The present study investigated whether bias in SUVR relative to DVR could be explained by factors such as underlying Aβ pathology and CBF (as measured by R_1). For both tracers, strong correlations were observed between SUVR and DVR, although SUVR overestimated DVR, and the magnitude of overestimation was positively associated with the degree of Aβ pathology. Furthermore, negative associations were observed between SUVR or $SUVR_{bias}$ and R_1, albeit not statistically significant.

For both tracers, the observed proportional overestimation of SUVR relative to DVR is in line with previous work using [11C]PiB, where similar results in terms of overestimation and proportionality of bias were observed when comparing SUVR with DVR (*15*). Carson and colleagues have mathematically demonstrated that target to reference tissue ratios (such as SUVR) overestimate true binding (such as DVR in the present work) at transient equilibrium conditions (*1*). In particular, the magnitude of overestimation depends on tracer kinetics, with slower kinetics corresponding to larger overestimation. The present findings are in line with this work, given that a larger overestimation was observed for these Fluorine-18 tracers, as compared with [11C]PiB, which has faster kinetics (*15*,*32*). In addition, Carson and colleagues showed increased overestimation for high binding regions, as a result of the longer time required to reach equilibrium and the lower tissue clearance rates (*1*), which is in agreement with the present findings. Together, these findings suggest that in cross-sectional analyses including participants with low amyloid burden, the expected bias in global SUVR will be small, although in early binding regions it might already be substantial. Furthermore, in longitudinal analyses, caution is warranted as this proportional bias can falsely inflate SUVR-based accumulation rates (*33*).

In mostly cognitively unimpaired individuals, there was no statistically significant effect of CBF on $SUVR_{bias}$, although a negative association was observed between SUVR or $SUVR_{bias}$ and CBF, as measured with R_1. These findings appear to be in disagreement with

a longitudinal [^{11}C]PiB study of van Berckel and colleagues, which reported decreases in global R_1 and concomitant decreases in global SUVR (*2*). The discrepancy with the present study is likely due to differences in the study populations included. In the present study, the vast majority of participants were cognitively unimpaired, while in the study of van Berckel et al., these findings were related only to patients with AD dementia. It is known that large changes in CBF primarily occur at later disease stages, which could explain the association between R_1 and SUVR in AD dementia patients in the study of van Berckel et al. and the lack of significant association in mainly cognitively unimpaired participants in the present study (*5*). Interestingly, most regions showed a negative association between R_1 and SUVR or $SUVR_{bias}$ in this study, as would be expected with the progression of AD over time (*5*,*34*,*35*). Furthermore, these findings demonstrate that changes in CBF are still small in mostly CU participants and that the effect on bias in SUVR is limited. However, this might not necessarily be the case when following these elderly individuals over time into advanced disease phases.

It should be noted, that there were some between-tracer differences in terms of the relationship between SUVR and DVR and the association between R_1 and SUVR or $SUVR_{bias}$. These differences may be partially explained by differences in tracer kinetics, related to tracer clearance, degree of non-specific binding, or adherence to reference tissue assumptions (*7*,*8*). In addition, differences in demographic factors such as the percentage of visually Aβ-positive participants, age, sex, or *APOE*-ε4 carriers, also exerted an effect on the observed relationship between SUVR and DVR, as illustrated by the significant effect of some of these variables on the GLM analyses (*15*). Furthermore, differences in these variables and the number of scanned participants per institute likely also explain the significant association between centre and SUVR or $SUVR_{bias}$ for some regions (supplementary Table 3). Last, it is important to note that the large difference in sample size between the two tracer cohorts possibly also played a role; the [^{18}F]flutemetamol cohort consisted of almost three times as many participants (N=90) compared with the [^{18}F]florbetaben cohort (N=31).

A limitation of the present work is that it investigated only cross-sectional relationships between SUVR, DVR, and R_1, as no longitudinal data were available. However, longitudinal data are currently being collected in these subjects. Thus, in the future, it would be interesting to determine the magnitude of change in R_1 over time in early AD stages and assess the relationship with bias in SUVR relative to DVR.

Conclusion

The present findings show that the degree of bias in SUVR relative to DVR in mostly cognitively unimpaired individuals is directly related to the extent of underlying Aβ pathology for both [^{18}F]flutemetamol and [^{18}F]florbetaben, which may be particularly problematic for longitudinal study designs. Furthermore, there were no statistically significant associations between R_1 and bias in SUVR relative to DVR, likely because changes in CBF are still limited in participants with early Aβ pathology.

Acknowledgements

The authors would like to thank all staff of the various centres for skilful acquisition of the scans.

References

1. Carson RE, Channing MA, Blasberg RG, et al. Comparison of Bolus and Infusion Methods for Receptor Quantitation: Application to [18 F]Cyclofoxy and Positron Emission Tomography. *Journal of Cerebral Blood Flow & Metabolism*. 1993;13:24-42.
2. Berckel van BNM, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. *J Nucl Med*. 2013;54:1570-1576.
3. Collij LE, Konijnenberg E, Reimand J, et al. Assessing Amyloid Pathology in Cognitively Normal Subjects Using ^{18} F-Flutemetamol PET: Comparing Visual Reads and Quantitative Methods. *J Nucl Med*. 2019;60:541-547.
4. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. *J Nucl Med*. 2017;58:1019-1024.
5. Jack CR, Knopman DS, Jagust WJ, et al. Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet Neurol*. 2013;12:207-216.
6. Golla SS, Verfaillie SC, Boellaard R, et al. Quantification of [18F]florbetapir: A test–retest tracer kinetic modelling study. *J Cereb Blood Flow Metab*. June 2018:0271678X18783628.
7. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18)F]flutemetamol. *Neuroimage*. 2015;121:184-192.
8. Becker GA, Ichise M, Barthel H, et al. PET quantification of 18F-florbetaben binding to β-amyloid deposits in human brains. *J Nucl Med*. 2013;54:723-731.
9. Pascoal TA, Therriault J, Mathotaarachchi S, et al. Topographical distribution of Aβ predicts progression to dementia in Aβ positive mild cognitive impairment. *Alzheimers Dement (Amst)*. 2020;12.
10. Sakr FA, Grothe MJ, Cavedo E, et al. Applicability of in vivo staging of regional amyloid burden in a cognitively normal cohort with subjective memory complaints: the INSIGHT-preAD study. *Alzheimer's Research & Therapy*. 2019;11:15.
11. Guo T, Landau SM, Jagust WJ, Alzheimer's Disease Neuroimaging Initiative. Detecting earlier stages of amyloid deposition using PET in cognitively normal elderly adults. *Neurology*. 2020;94:e1512-e1524.
12. Collij LE, Heeman F, Salvadó G, et al. Multi-tracer model for staging cortical amyloid deposition using PET imaging. *Neurology*. July 2020:10.1212/WNL.0000000000010256.
13. Binnewijzend MAA, Benedictus MR, Kuijer JPA, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. *Eur Radiol*. 2016;26:506-514.
14. van Assema DME, Lubberink M, Boellaard R, et al. P-Glycoprotein Function at the Blood–Brain Barrier: Effects of Age and Gender. *Mol Imaging Biol*. 2012;14:771-776.

15. Lopes Alves I, Collij LE, Altomare D, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. *Alzheimer's Dement*. April 2020.
16. Vizamyl.https://www.gehealthcare.co.uk/en/products/categories/nuclear_imaging_agents/'vizamyl.
17. Heeman F, Yaqub M, Lopes Alves I, et al. Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. *EJNMMI Research*. 2019;9:32.
18. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*. 2010;29:196-205.
19. Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M. Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. *Front Neuroinform*. 2014;7.
20. Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. LEAP: Learning embeddings for atlas propagation. *NeuroImage*. 2010;49:1316-1325.
21. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. *J Cereb Blood Flow Metab*. 2002;22:1440-1452.
22. Boellaard R, Yaqub M, Lubberink M, Lammertsma A. PPET: A software tool for kinetic and parametric analyses of dynamic PET studies. *NeuroImage*. 2006;Supplement 2:T62.
23. Heeman F, Yaqub M, Hendriks J, et al. Parametric imaging of dual-time window [18F] flutemetamol and [18F]florbetaben studies. *NeuroImage*. 2021;234:117953.
24. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. *NeuroImage*. 1997;6:279-287.
25. Peretti DE, Reesink FE, Doorduin J, et al. Optimization of the k2′ Parameter Estimation for the Pharmacokinetic Modeling of Dynamic PiB PET Scans Using SRTM2. *Front Phys*. 2019;7.
26. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia*. 2015;11:1-15.e4.
27. Insel PS, Mormino EC, Aisen PS, Thompson WK, Donohue MC. Neuroanatomical spread of amyloid β and tau in Alzheimer's disease: implications for primary prevention. *Brain Commun*. 2020;2:fcaa007.
28. Mattsson N, Palmqvist S, Stomrud E, Vogel J, Hansson O. Staging β-Amyloid Pathology With Amyloid Positron Emission Tomography. *JAMA Neurol*. July 2019.
29. Grecchi E, Foley C, Gispert JD, Wolz R. P3-434: Centiloid Pet Suvr Analysis Using the Supratentorial White Matter as Reference Region. *Alzheimer's & Dementia*. 2018;14:P1278-P1278.

30. Martin Bland J, Altman DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet*. 1986;327:307-310.
31. Mukaka M. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med J*. 2012;24:69-71.
32. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Annals of Neurology*. 2004;55:306-319.
33. Lopes Alves I, Heeman F, Collij LE, et al. Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging. *Alzheimers Res Ther*. 2021;13:82.
34. Duan W, Sehrawat P, Balachandrasekaran A, et al. Cerebral Blood Flow Is Associated with Diagnostic Class and Cognitive Decline in Alzheimer's Disease. *J Alzheimers Dis*. 2020;76:1103-1120.
35. Matsuda H. Cerebral blood flow and metabolic abnormalities in Alzheimer's disease. *Ann Nucl Med*. 2001;15:85.

Supplementary materials

Supplementary Table 1. Acquisition methods per centre

	UEDIN	BBRC	Amsterdam UMC, VUmc	UNIGE
Scanner manufacturer	Siemens	Siemens	Philips	Siemens
Scanner model	Biograph mMR	Biograph64 mCT	Ingenuity TF PET/MR	Biograph128 mCT
Pixel Spacing	1.043*1.043	1.0182*1.018	2.000*2.000	1.018*1.018
Reconstruction method	OP-OSEM4i21s	OSEM3D 4i24s	LOR-RAMLA	OSEM3D 4i24s
Attenuation correction	MR-based, HiRes	CT-based	MR-based	CT-based
Scatter correction	Model based, relative, single scatter simulation	Model-based, relative, single scatter simulation	SS-SIMUL	Model-based, relative, single scatter simulation

Amsterdam UMC, VUmc = Amsterdam University Medical Centre, VUmc BBRC = Barcelona Beta Brain Research Center, UNIGE= University of Geneva, UEDI*N*= The University of Edinburgh. OSEM: ordered-subset expectation maximization, LOR-RAMLA: standard line-of-response–based row-action maximum-likelihood algorithm

Supplementary Table 2a. Relationship between (bias in) SUVR and independent variables for [^{18}F]flutemetamol

Precuneus	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.84#	1.66 - 2.03	33.49#	21.49 – 45.48
R_1	-0.20	-0.43 - 0.03	-8.96	-24.02 – 6.09
Age	0.00	0.00 - 0.01	0.23	-0.01 – 0.47
Sex	-0.01	-0.06 - 0.03	-0.38	-3.12 – 2.35
***APOE*-ε4**	0.05*	-0.01 - 0.09	3.30*	-0.55 – 6.06
BBRC	-0.05	-0.14 - 0.04	-3.50	-9.46 – 2.46
Amsterdam UMC	-0.04	-0.14 - 0.05	-2.14	-8.41 – 4.13
UNIGE	-0.03	-0.15 - 0.10	-3.07	-11.23 – 5.09

Females, *APOE*-ε4 non-carriers and centre UEDIN were used as reference groups. *p<0.05, **p<0.01, #p<0.001

Supplementary Table 2b. Relationship between (bias in) SUVR and independent variables for [^{18}F]flutemetamol

PCC	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.60#	1.42 - 1.79	13.54*	3.08 – 23.99
R_1	-0.20	-0.44 - 0.04	-9.99	-23.29 – 3.32
Age	0.00	0.00 - 0.01	0.24*	0.03 – 0.45
Sex	-0.01	-0.06 - 0.03	-0.77	-3.12 – 1.57
***APOE*-ε4**	0.03	-0.01 - 0.07	1.35	-0.97 – 3.66
BBRC	0.00	-0.09 - 0.09	-0.70	-5.72 – 4.32
Amsterdam UMC	0.03	-0.07 - 0.12	1.41	-3.82 – 6.63
UNIGE	0.06	-0.07 - 0.18	2.77	-4.20 – 9.75

Females, *APOE*-ε4 non-carriers and centre UEDIN were used as reference groups. PCC: posterior cingulate cortex, * $p<0.05$, ** $p<0.01$, # $p<0.001$

Supplementary Table 2c. Relationship between (bias in) SUVR and independent variables for [^{18}F]flutemetamol

OFC	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.60#	1.42 - 1.79	25.06**	8.99 – 41.12
R_1	0.02	-0.31 - 0.35	-6.77	-21.28 – 34.82
Age	0.00	0.00 - 0.01	0.34	-0.04 – 0.71
Sex	-0.02	-0.06 - 0.03	-1.84	-5.79 – 2.11
***APOE*-ε4**	0.01	-0.04 - 0.06	1.10	-2.90 – 5.11
BBRC	0.03	-0.07 - 0.14	2.99	-5.79 – 11.77
Amsterdam UMC	0.06	-0.05 - 0.17	5.99	-3.37 – 15.35
UNIGE	0.16*	0.01 - 0.30	-12.17	-0.16 – 24.50

Females, *APOE*-ε4 non-carriers and centre UEDIN were used as reference groups. OFG: orbitofrontal gyrus, *$p<0.05$, **$p<0.01$, #$p<0.001$

Supplementary Table 2d. Relationship between (bias in) SUVR and independent variables for [^{18}F]florbetaben

Precuneus	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.41#	1.27 - 1.55	17.94#	5.78 - 30.11
R_1	-0.05	-0.35 - 0.25	-1.83	-27.45 - 23.80
Age	0.00	0.00 - 0.01	0.32	0.00 - 0.64
Sex	-0.06	-0.11 - 0.00	-4.54	-9.42 - 0.35
APOE-ε4	0.02	-0.03 - 0.08	2.00	-2.77 - 6.78

Females and *APOE*-ε4 non-carriers were used as reference groups. *$p<0.05$, **$p<0.01$, #$p<0.001$

Supplementary Table 2e. Relationship between (bias in) SUVR and independent variables for [^{18}F]florbetaben

PCC	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.43#	1.21 - 1.65	15.14	-3.79 – 34.07
R_1	-0.21	-0.59 - 0.17	-17.63	-50.15 – 14.90
Age	0.00	0.00 - 0.01	0.32	-0.07 – 0.71
Sex	-0.05	-0.12 - 0.01	-4.57	-10.55 – 1.41
***APOE*-ε4**	0.06	-0.01 - 0.12	4.75	-0.94 – 10.43

Females and *APOE*-ε4 non-carriers were used as reference groups. PCC: posterior cingulate cortex, *$p<0.05$, **$p<0.01$, #$p<0.001$

Supplementary Table 2f. Relationship between (bias in) SUVR and independent variables for [^{18}F]florbetaben

OFC	SUVR		$SUVR_{bias}$	
	Coefficient estimate	95% Confidence interval	Coefficient estimate	95% Confidence interval
DVR	1.37#	1.15 - 1.59	16.74	-1.87 – 35.36
R_1	0.01	-0.46 - 0.47	1.85	-37.01 – 40.70
Age	0.00	-0.01 - 0.00	-0.13	-0.56 – 0.31
Sex	-0.05	-0.13 - 0.03	-4.36	-10.92 – 2.20
***APOE*-ε4**	-0.01	-0.08 - 0.06	-0.45	-6.60 – 5.69

Females and *APOE*-ε4 non-carriers were used as reference groups. OFG: orbitofrontal gyrus, *$p<0.05$, **$p<0.01$, #$p<0.001$

Supplementary Table 3. Demographic differences between centres

Centre	# Participants	Age	Females (%)	MMSE	*APOE*-ε4 Carriers (%)	DVR	R_1
UEDIN	5	62.6	40.0	29.2	60.0	1.16	1.07
BBRC	29	64.7	51.7	29.0	27.6	1.15	1.03
Amsterdam UMC	51	68.7	54.9	29.0	45.1	1.26	0.97
UNIGE	5	69.6	80.0	27.4	60.0	1.40	1.03

8.

[^{11}C]PiB amyloid quantification: effect of reference region selection

Fiona Heeman, Janine Hendriks, Isadora Lopes Alves, Rik Ossenkoppele, Nelleke Tolboom, Bart N.M. van Berckel, Adriaan A. Lammertsma, Maqsood Yaqub
On behalf of the AMYPAD Consortium

***As published in EJNMMI Research,* 2020; 10:123**

Abstract

Background: The standard reference region (RR) for amyloid-beta (Aβ) PET studies is the cerebellar grey matter (GMCB), while alternative RRs have mostly been utilised without prior validation against the gold standard. This study compared five commonly used RRs to gold standard plasma-input-based quantification using the GMCB.

Methods: 13 subjects from a test-retest (TRT) study and 30 from a longitudinal study were retrospectively included (total: 17 Alzheimer's disease, 13 mild cognitive impairment, 13 controls). Dynamic [^{11}C]PiB PET (90 min) and T1-weighted MR scans were co-registered and time-activity curves were extracted for cortical target regions and the following RRs: GMCB, whole cerebellum (WCB), white matter brainstem/pons (WMBS), whole brainstem (WBS) and eroded subcortical white matter (WMES). A two-tissue reversible plasma input model (2T4k_V_b) with GMCB as RR, reference Logan and the simplified reference tissue model were used to derive distribution volume ratios (DVRs), and standardized uptake value (SUV) ratios were calculated for 40-60 min and 60-90 min intervals. Parameter variability was evaluated using TRT scans, and correlations and agreements with the gold standard (DVR from 2T4k_V_b with GMCB RR) were also assessed. Next, longitudinal changes in SUVs (both intervals) were assessed for each RR. Finally, the ability to discriminate between visually Aβ-positive and Aβ-negative scans was assessed.

Results: All RRs yielded stable TRT performance (max 5.1% variability), with WCB consistently showing lower variability. All approaches were able to discriminate between Aβ-positive and Aβ-negative scans, with highest effect sizes obtained for GMCB (range:-0.9 to -0.7), followed by WCB (range:-0.8 to -0.6). Furthermore, all approaches provided good correlations with the gold standard ($r \geq 0.78$), while the highest bias (as assessed by the regression slope) was observed using WMES (range slope: 0.52-0.67), followed by WBS (range slope: 0.58-0.92) and WMBS (range slope: 0.62-0.91). Finally, RR SUVs were stable across a period of 2.6 years for all except WBS and WMBS RRs (60-90 min interval).

Conclusions: GMCB and WCB are considered the best RRs for quantifying amyloid burden using [^{11}C]PiB PET.

Keywords: [^{11}C]PiB, Alzheimer's disease, amyloid PET, quantification, reference regions

Background

Amyloid-beta accumulation (Aβ) in the brain is a pathological hallmark of Alzheimer's disease (AD) and can be measured *in vivo* using positron emission tomography (PET) (*1*,*2*). One of the first amyloid PET tracers is Pittsburgh compound B ([^{11}C]PiB), which binds with high specificity to fibrillar Aβ deposits (*3*,*4*). Both static and dynamic PET image acquisition protocols have been used, where the first is often preferred for routine and multi-centre studies due to its short duration and relatively simple processing. However, a static scan only provides a semi-quantitative measure of amyloid load, which can be affected by confounders (*5*–*8*). Therefore, performing dynamic acquisitions and full quantification using kinetic modelling may be required for assessing subtle changes in amyloid load, which is of particular importance in longitudinal studies where other physiological parameters may change, thereby introducing bias (*5*). In general, a disadvantage of such a protocol is the need for arterial sampling, which is logistically challenging, requires specially trained staff and dedicated equipment, and is particularly burdensome to the patient. A possible alternative to the use of arterial sampling is a reference tissue approach (*9*). Reference tissue approaches rely on the assumption that a region devoid of specific binding, but otherwise having similar tissue characteristics as the target region of interest, is available (=reference region), providing an indirect input function and circumventing the need for arterial sampling (*10*,*11*).

In case of imaging Aβ deposits in AD using [^{11}C]PiB, the cerebellar grey matter (GMCB) meets the assumptions of a reference region in nearly all patients, and it has been validated against the plasma input approach (*6*,*12*). Only in rare familial forms and advanced stages of AD, this region might become compromised with Aβ plaques (*13*,*14*). In addition, accurate segmentation of this region can be challenging, and may be hampered by truncation of the field of view in the lower portion of the brain. In recent years, several reports have proposed alternative reference regions, either aiming to overcome these issues, or aiming to improve effect sizes when measuring Aβ changes over time (*15*–*17*). However, these alternative reference regions do not necessarily meet all requirements for a suitable reference tissue, such as having the same tissue characteristics as the target tissue or showing longitudinal stability and similar behaviour across diagnostic groups (*15*,*16*). One such region often used for amyloid quantification is whole cerebellum (*16*,*17*). Alternatively, reference tissues predominantly consisting of white matter, such as brainstem/pons or eroded subcortical white matter (centrum semiovale) have been proposed, in particular, for longitudinal amyloid quantification (*17*,*18*). However, age-

related changes have been reported in the non-specific tracer retention of white matter regions, possibly compromising their use for longitudinal amyloid quantification (*19*).

To date, the impact of using alternative reference regions (RRs) on amyloid quantification has mainly been evaluated for semi-quantitative parameters (*15–17*). Most alternative RRs have not been validated against the gold standard, i.e. full quantification with metabolite corrected plasma input curves or full quantification using a validated reference region. Therefore, the present work focussed on the widely used [^{11}C]PiB amyloid PET tracer and evaluated the use of the validated cerebellar grey matter as well as four alternative reference regions: whole cerebellum, white matter brainstem/pons, whole brainstem and eroded subcortical white matter. The performance of these regions was evaluated for both semi- and fully quantitative analysis in a test-retest (TRT) and longitudinal setting in terms of precision with respect to TRT variability, accuracy compared with the gold standard, stability over time (in case of the standardized uptake value, SUV), power for group discrimination and detecting physiologically plausible, longitudinal accumulation processes.

Materials & Methods

Subjects

Clinical data of 43 participants belonging to two different studies, both conducted within the Amsterdam UMC, location VUmc, were included retrospectively (*20,21*). Thirteen subjects (6 cognitively unimpaired (CU), 1 mild cognitive impaired (MCI), 6 AD) were part of a TRT study and underwent arterial sampling, as described in detail by Tolboom et al. (*21*). The other 30 subjects (11 CU, 12 MCI, 7 AD) were part of a longitudinal study as described by Ossenkoppele et al. (*20*). In brief, all subjects received standard dementia screening for diagnostic purposes and amyloid PET scans were assessed visually (positive or negative) (*21,22*). Before enrolment, all participants provided written informed consent and the Medical Ethics Review Committee of the Amsterdam UMC, location VUmc, had approved both studies.

Image acquisition

All subjects from the TRT study underwent a structural T1-weighted MR scan on a 1.5T Siemens Sonata scanner (MPRAGE: matrix size 256×256 and 160 slices, voxel size 1.0×1.0×1.5 mm, echo time = 3.97 ms, repetition time = 2.700 ms, inversion time = 950 ms, flip angle 8°) and a test and same-day retest dynamic [^{11}C]PiB PET scan (except

for one subject) on a Siemens ECAT EXACT HR+ scanner (*21*). All participants first received a 10-minutes transmission scan for photon attenuation correction, followed by an intravenous [^{11}C]PiB injection and simultaneously starting a 90 minutes dynamic PET scan (*21*). Arterial blood was monitored continuously for the first 60 minutes using an online detection system and additional manual samples were drawn for calibration, to determine plasma to whole-blood ratios, and to measure plasma parent and metabolite fractions (*21*). For seven subjects, arterial blood data were not available or not of sufficient quality for at least one of the scans. In addition, for one subject, the second scan was not used due to severe motion between PET frames. Consequently, a total of *N*=6 test scans and *N*=5 retest scans with plasma input data were available. With respect to the longitudinal study, subjects also underwent similar T1-weighted MR and dynamic [^{11}C]PiB PET scans at baseline, and follow-up (same scanners), 30.3±5.4 (range: 23-48) months later, but no arterial blood was sampled (*20*).

Image processing

First, structural T1-weighted MR images were co-registered to their corresponding PET image. Next, PVE-lab software was used to segment grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF), as well as to delineate volumes of interest (VOIs) based on the Hammers atlas (*23,24*). The following grey matter regions were used as target regions: medial and lateral anterior temporal lobe, posterior temporal lobe, superior, middle and inferior temporal gyrus, fusiform gyrus, parahippocampal and ambient gyrus, anterior and posterior cingulate gyrus, middle and orbitofrontal gyrus, gyrus rectus, inferior and superior frontal gyrus, pre- and post-central gyrus, superior parietal gyrus and the (infero)lateral remainder of the parietal lobe. In addition, a composite global cortical region was generated as the volume-weighted average across all target regions. The RRs included GMCB, whole cerebellum (WCB), white matter brainstem/pons (WMBS), whole brainstem (WBS) and the eroded subcortical white matter (WMES). The WMES was obtained by eroding the subject's whole brain WM segmentation (using the *imerode* function in MATLAB) and manually removing cerebellar and brainstem white matter. Corresponding time-activity curves (TACs) were obtained by superimposing VOIs on the dynamic PET scan.

Kinetic analysis

Only for scans where arterial plasma input data were available, the reversible two-tissue compartment model with four rate constants and additional blood volume fraction parameter (2T4k_V_b) was used to estimate the volume of distribution (V_T). Volume

of distribution ratios (DVR_{2T4k_Vb} = V_T target / V_T reference) were calculated indirectly by using the validated GMCB as RR (here called: DVR_{PI_GMCB}) (=gold standard) (*6*,*12*).

For all scans, reference Logan (RLogan) was used to estimate DVR (DVR_{RLOGAN}). The implementation did not require fixing k_2' (as per Eq. 7 from Logan et al. (*25*) and a linearization time (t^*) of 50 min p.i. was used (*6*,*25*). In addition, the simplified reference tissue model (SRTM) was used to estimate binding potential (BP_{ND}) with parameter fit boundaries optimised per RR (see Supplementary Table 1), and BP_{ND}+1 (=DVR) was calculated for comparison (*10*). Finally, standardized uptake value ratios (SUVr) were calculated for two frequently used acquisition windows (40-60 and 60-90 min p.i., $SUVr_{40-60}$ and $SUVr_{60-90}$, respectively) (*6*,*12*). For each reference tissue method all RRs mentioned above were used.

Statistical analysis

Statistical analyses were performed in IBM SPSS Statistics for Windows Version 24.0 (IBM Corp. Armonk New York U.S.A), GraphPad Prism for Windows Version 7.04 (La Jolla California U.S.A.), and Origin Version 2019b (OriginLab Corporation, Northampton, Massachusetts, USA). For each reference region and method, regional outliers were defined based on the median absolute deviation (MAD3) criterion assuming a non-normal distribution (*26*). This resulted in a total of 36 values, across all subjects (from the 2T4k_V_b and SRTM models) being excluded from further analyses (see supplementary materials for details). Differences in age and score on the Mini-Mental State Examination (MMSE) between diagnostic groups were assessed using non-parametric Kruskal-Wallis and post-hoc Mann-Whitney *U* tests, while differences in the proportion of males and females were tested with chi-square tests. As the TRT cohort consisted of only one MCI subject, this subject was not used for comparison.

Test-retest cohort

First, using the composite global cortical value, relative test-retest variability was calculated per RR and method according to Eq. 1, where the estimate of global cortical amyloid load (DVR or SUVr) of the test scan is denoted as T and for the retest scan as R.

$$TrT\ variability\ (\%) = \frac{|T - R|}{0.5 \cdot |T + R|} \cdot 100 \qquad (1)$$

Second, based on results obtained from the test scans (N=6), agreement between regional quantification (for all RRs and reference tissue methods) and the gold standard

(DVR_{PI_GMCB}) was assessed using Bland-Altman (BA) analyses (*27*). Next, linear regression analysis of the data points in the BA plots was used to assess whether (and to what extent) bias was dependent on underlying amyloid burden. Finally, correlations, slopes and intercepts between DVR_{PI_GMCB} and the corresponding parameter of interest derived from each of the RRs and methods were calculated using linear regression analysis.

Longitudinal cohort

A sub-set of subjects (N=18) had information available on injected dose and patient weight, for which SUV TACs were calculated for all RRs. In addition, mean SUVs were calculated for all RRs and both acquisition windows (40-60 min and 60-90 min p.i.). The shape of the SUV TACs were assessed in the baseline scans, and the stability of the RRs over time was assessed using paired t-tests with Bonferroni correction. Follow-up time was standardized to the average follow-up time across subjects (2.6 years) to account for between-subject differences.

Finally, as an exploratory analysis, the annual percentage change in the composite global cortical value was calculated per individual and for each of the RRs (according to Eq. 2)

$$Annual\ percentage\ change = \frac{FU - BL}{Years} \cdot \frac{100}{BL} \qquad (2)$$

With the parameter at follow-up scan as *FU*, at baseline scan as *BL* and *Years* stands for the number of years since baseline scan. These values were plotted against the baseline parameter and the relationship was assessed by fitting linear and quadratic models through the data. These models were chosen based on previous literature and the known dose-response relationship of binding (*1*,*28*,*29*), where the hypothesis is that amyloid burden measured with PET plateaus at later stages of the disease (*1*). Goodness of fit was assessed using the Akaike Information Criterion (AIC) (*30*).

Discriminative ability reference regions

For the global cortical parameter of interest, derived using each of the methods and RRs, the ability to discriminate between visually Aβ-positive and-Aβ negative scans was assessed using Mann-Whitney *U* tests with Bonferroni correction (using scans with stable longitudinal visual assessment: N=80). In addition, the Hodges-Lehmann estimate of the median difference was used as measure of the effect size (*31*).

Results

Subjects

Demographics are presented in Table 1. As expected, CU subjects had higher MMSE scores (i.e. better global cognition) than AD subjects in both TRT (p=0.003) and longitudinal (p=0.001) studies. In addition, in the longitudinal study, higher MMSE scores were observed for CU compared with MCI subjects (p=0.005), as well as a trend towards higher MMSE scores for MCI compared with AD subjects (p=0.083). There were no differences with respect to age and sex.

Table 1. Subject demographics

TRT	**CU (*N*=6)**	**MCI (*N*=1)**	**AD (*N*=6)**
Age	64.3 ±5.7	71.0	61.0 ±3.0
Females (%)	50%	100%	17%
MMSE	29.7 ±0.5	28.0	20.7 ±2.0
Longitudinal	**CU (*N*=11)**	**MCI (*N*=12)**	**AD (*N*=7)**
Age	66.4 ± 7.3	67.4 ±6.7	60.4 ± 5.4
Females (%)	27%	33%	14%
MMSE	29.4 ± 0.5	27.2 ±2.5	25.3 ±2.3

Values are depicted as M±SD

Test-retest cohort

Test-retest variability

The maximum TRT variability across regions and methods was 5.1%, with lowest TRT variability observed for WCB across methods (Table 2). Across RRs, RLogan showed least variability overall, while $SUVr_{40\text{-}60}$ showed less variability than $SUVr_{60\text{-}90}$ (Table 2). Furthermore, the Bland-Altman analyses showed that for all RRs and methods, variability was most pronounced at low SUVR and DVR values (Figure 1) and highest for the WMES (Supplementary Table 2a).

Table 2. Relative test-retest variability across reference regions and methods

	DVR_{RLOGAN}	**DVR_{SRTM}**	**$SUVr_{40\text{-}60}$**	**$SUVr_{60\text{-}90}$**
GM Cerebellum	2.8	2.9	3.5	5.1
Whole Cerebellum	1.4	2.0	2.2	2.8
WM Brainstem /Pons	2.4	3.3	2.3	3.7
Whole Brainstem	2.1	3.8	2.2	3.1
Subcortical Eroded WM	2.4	2.7	3.7	3.9

All values are % TRT variability of global cortical averages for N=12

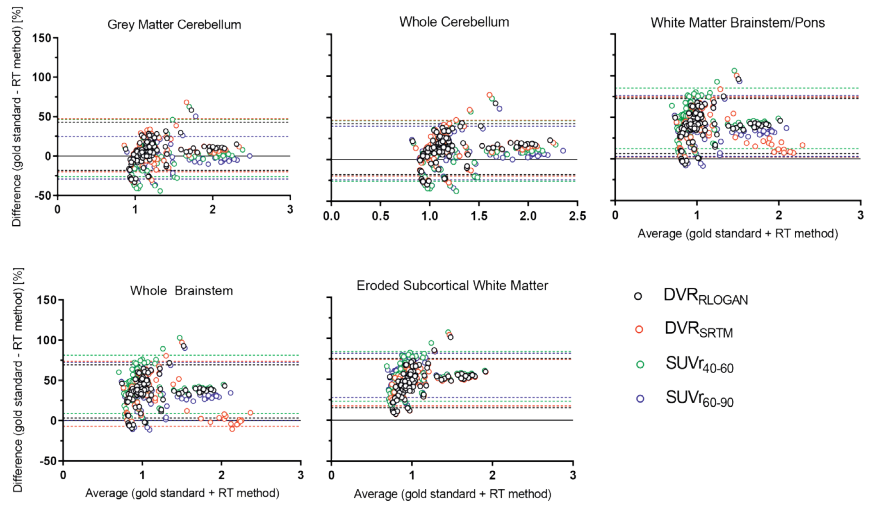


Figure 1. Bland-Altman: agreement with the gold standard for each reference region
Bland-Altman plot for each of the reference regions, showing the performance of all methods. RT: refers to the reference tissue method that is being compared.

Agreement with gold standard

Across methods, all RRs showed a strong correlation (r≥0.78) with the gold standard, DVR_{PI_GMCB} (Table 3). Furthermore, GMCB and WCB RRs showed the smallest bias across methods as indicated by the regression slopes (Table 3, range: 0.85-1.12, 0.81-1.05, respectively) and WMES the worst (Table 3, range: 0.57-0.67) and shown by the Bland-Altman analysis (Figure 1 and Supplementary Table 2a). However, using RRs that contained white matter resulted in an underestimation compared with DVR_{PI_GMCB} for all parameters except SUVr's calculated using the WCB (Table 3 and Figure 1). In addition, the bias introduced by using WMES RR showed the strongest dependency on the underlying amyloid burden (Figure 1 and Supplementary Table 2b). Finally, across methods, $SUVr_{60\text{-}90}$ showed a better correlation with DVR_{PI_GMCB} than $SUVr_{40\text{-}60}$ (Table 3).

Table 3. Test-retest cohort: correlations between reference tissue methods with varying RRs and the gold standard: DVR_{PI_GMCB}

Reference Region		DVR_{RLOGAN}	DVR_{SRTM}	$SUVr_{40-60}$	$SUVr_{60-90}$
GM Cerebellum	*r*	0.88	0.85	0.81	0.89
	Slope	0.85	0.85	1.04	1.12
	Intercept	0.14	0.18	-0.03	-0.10
Whole Cerebellum	*r*	0.85	0.81	0.77	0.84
	Slope	0.81	0.84	1.00	1.05
	Intercept	0.09	0.03	-0.11	-0.15
WM Brainstem/ Pons	*r*	0.81	0.84	0.78	0.80
	Slope	0.73	0.62	0.81	0.91
	Intercept	-0.07	0.04	-0.24	-0.30
Whole Brainstem	*r*	0.81	0.83	0.79	0.80
	Slope	0.74	0.58	0.83	0.92
Subcortical	Intercept	-0.06	0.11	-0.24	-0.29
	r	0.83	0.86	0.82	0.86
Eroded WM	Slope	0.57	0.52	0.67	0.63
	Intercept	0.07	0.13	-0.10	-0.08

Values are shown for each of the methods and correspond to the linear regression analysis

Longitudinal cohort

SUV reference region TACs and stability over time

SUV TACs of the five RRs are depicted in Figure 2, illustrating that WCB and GMCB, as well as WBS and WMBS showed a very similar shape. Furthermore, cerebellar RRs showed the steepest decline in uptake over time, followed by brainstem RRs and cerebellar and WMES RR TACs differed most. With respect to the stability of longitudinal SUV uptake (60-90 min p.i.), significant decreases (after Bonferroni correction) between baseline and follow-up SUV measurements were only present for WBS and WMBS (p=0.004, p=0.003) and a trend level decrease was observed for WCB (p=0.006). With respect to the early acquisition window (40-60 min p.i.), no significant differences were present, although the strongest trend was observed for WBS and WMBS.

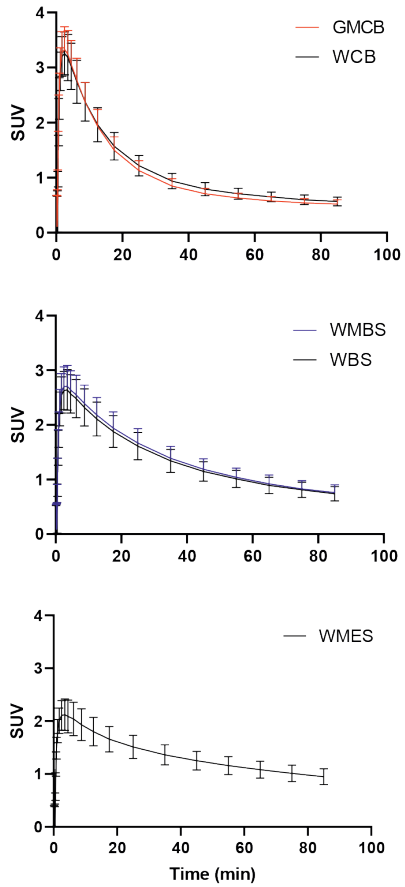


Figure 2. SUV TACs for all reference regions
Standardized uptake value time-activity curves (corrected for weight and injected dose)

Annual change and baseline amyloid load

Across methods, the relationship between annual percentage change and baseline amyloid load as obtained by GMCB and WCB was best described by a quadratic relationship (Figure 3) (ΔAIC GMCB RLogan: 8.0, SRTM: 6.4, $SUVr_{40-60}$: 3.6, $SUVr_{60-90}$: 2.4 and ΔAIC WCB RLogan: 4.2, SRTM: 12.3, $SUVr_{40-60}$: 2.9, $SUVr_{60-90}$: 2.0). In contrast, for WMBS and WBS the relationship was best described by a quadratic model for SRTM (ΔAIC: 8.5 and 12.2, respectively) and $SUVr_{40-60}$ (ΔAIC: 3.3 and 3.8, respectively) and by a linear model for RLogan (ΔAIC: 1.6 and 1.5, respectively) and $SUVr_{60-90}$ (ΔAIC: 2.4 and 2.5, respectively) (Figure 3). Finally, with respect to WMES, the relationship was best

described by a linear model for all methods (ΔAIC RLogan: 0.8, SRTM: 1.2 $SUVr_{40-60}$: 0.0, $SUVr_{60-90}$: 1.8).

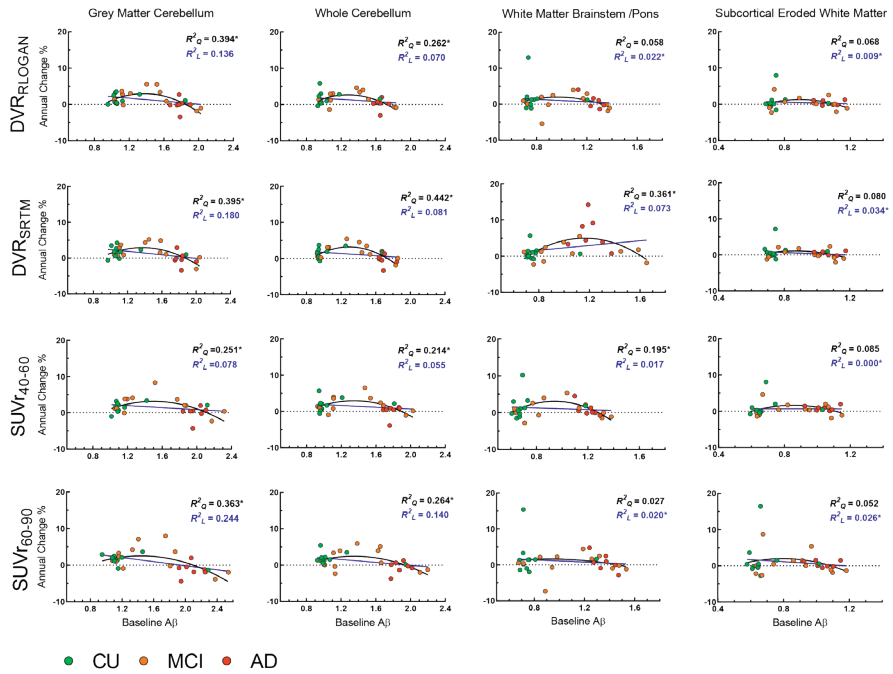


Figure 3. Baseline Aβ vs. annual percentage change across reference regions and methods
The asterisk indicates the model that was preferred by the AIC.

Scans from both cohorts

Discriminative ability reference regions
All parameters of interest derived using each of the RRs (and methods) were able to discriminate between Aβ-positive and Aβ-negative (p<0.001) scans (Supplementary Figure 1). Highest effect sizes were obtained for GMCB (range:-0.9 -0.7), followed by the WCB RR (range:-0.8 -0.6) and lowest effect sizes for WMES (-0.4) (Supplementary Table 3).

Discussion

In the present [11C]PiB study, the performance of five reference regions was evaluated. All reference regions yielded relatively small test-retest variability and showed good correlations with the gold standard DVR_{PI_GMCB}. However, largest bias, as shown by the regression slopes and BA analyses, was observed for white matter-based RRs. In addition, the choice of reference region did not impact the ability to differentiate between Aβ-positive and negative scans, but the largest effect sizes were obtained for GMCB and WCB. Furthermore, the longitudinal study showed that SUV changed over time for both WBS and WMBS RRs, but only when using the late acquisition window (60-90min). Finally, the relationship between baseline amyloid and Aβ accumulation was best described by a quadratic model, as expected, for GMCB and WCB.

While the maximum TRT variability was 5.1% across methods, the WCB RR showed consistently lower variability (Table 2). This may be related to the fact that this region is less prone to segmentation errors than for example GMCB and has more counts compared with the brainstem as a result of its larger volume. In addition, WCB may also outperform WMES in terms of TRT variability because the latter showed bias that was more dependent on the underlying amyloid burden (Figure 1 and Supplementary Table 2b). Finally, all regional parameters of interest, derived using all methods and RRs, showed good correlations (r≥0.78) with regional DVR_{PI_GMCB} (Table 3).

Using GMCB and WCB as RR yielded, as expected, least bias as compared with the gold standard (as shown by the linear regression: Table 3 and BA analysis: Figure 1 and Supplementary Table 2a). RRs that primarily contained white matter showed substantial underestimation compared with values obtained by the plasma input model, except for WCB, were this underestimation was only observed for RLogan and SRTM (Table 3). This underestimation is likely a result of both the relatively high uptake in white compared with grey matter and the different kinetics in this tissue compared with other RRs as illustrated by Figure 2. Furthermore, given that the two cerebellar as well as the two brainstem RR SUV TACs were very similar in shape, relatively small differences in performance with respect to precision and accuracy were expected. These findings also indicate that the effect of choice of tissue of the RR on quantification is smaller than the effect of using a different anatomical RR. Furthermore, for WMES RR, bias (as shown by the BA analysis) was most dependent on the underlying amyloid burden (Supplementary

Table 2b). Therefore, using WMES for normalisation purposes could be problematic, in particular for analysing regions or subjects spanning the AD continuum.

The longitudinal results showed significant decreases in WBS and WMBS SUV only for the late (60-90 min.) acquisition window. However, a similar trend (although not significant) was present for the early (40-60 min) acquisition window. This finding might be related to the fact that SUV does not take flow changes into account (*20*). As such, using WBS or WMBS for normalisation purposes, may result in an overestimation of the true Aβ load and this would be particularly problematic for longitudinal Aβ quantification (*19*). In fact, effects of these confounding factors may explain why some studies have reported increased power for detecting longitudinal changes or larger between group differences in rates of Aβ change, using white matter RRs (*32*,*33*). Moreover, decreases in white matter SUV also may explain the lower pons and WMES SUVR values for groups of increasing disease severity (using GMCB as RR), as previously reported by Tryputsen and colleagues (*34*), although the authors themselves provide a different explanation by suggesting it could be due to increasing GMCB Aβ load. Ideally, one would have used V_T for assessing the stability of RRs over time, but this was not possible as these subjects did not undergo arterial sampling.

Furthermore, results showed that although all RRs were able to discriminate between Aβ-positive and Aβ-negative scans, GMCB and WCB provided the highest effect sizes, while WMES provided the poorest results. Therefore, GMCB and WCB would be preferred for detecting more subtle between-group differences. These findings partially differ from some previous reports, likely due to differences in study population, study design or criteria used for defining the optimal RR. For example, some studies reported highest effect sizes for GMCB and pons or for WMES and pons when discriminating between diagnostic groups, which only partly agrees with the present results when discriminating between Aβ-positive and Aβ-negative scans (*35*,*36*). Moreover, the high effect sizes reported for WM RRs could also be related to the effect of confounding factors, as discussed above. It should be noted, however, that these results belong to a group-classification analysis, hence they cannot be compared directly with findings from studies assessing the statistical power for detecting longitudinal changes in Aβ burden, that employ a within-subject design (*32*,*34*). Finally, differences in the criteria used for identifying the optimal RR can have a significant impact on outcome. For example, while Schwarz and colleagues exclusively focused on longitudinal criteria to recommend a combination of voxels from supratentorial white matter and whole cerebellum as RR,

the present study used a combination of criteria based on a comparison against the gold standard, test-retest variability and longitudinal performance (*33*).

In the present study, an inverted u-shaped relationship between baseline amyloid load and Aβ accumulation was observed only for GMCB and WCB RRs. This pattern has been reported previously (*37*–*39*), and is in line with the known sigmoidal dose response relationship of binding (*29*). It should be noted that this was only an exploratory analysis, and further studies are needed to explore this relationship and possible between group differences in Aβ accumulation (e.g. by diagnostic or Aβ status) in a larger dataset.

Taken together, the present results suggest that GMCB and WCB are suitable RRs with respect to analysing [11C]PiB scans. Overall, accuracy as compared with the gold standard was higher using GMCB, while precision (as assessed by measurement variability and dependency of the bias on underlying Aβ burden) was more favourable using WCB. Therefore, in cross-sectional studies one might prefer GMCB, as it more closely adheres to the "truth", while in longitudinal studies, where stability of results outweighs a small bias, WCB would be preferred. Finally, it is important to note that the results of the present study relate to [11C]PiB and are not necessarily translatable to other tracers. As shown previously by Villemagne and colleagues for SUVr, the most stable RR may differ per tracer (*40*), and this finding was supported by studies using both [18F]florbetaben and [18F]florbetapir (*16*,*41*). These between tracer discrepancies may be the result of differences in non-specific binding in the reference region (as compared with the F-18 labelled tracers) or violations of the reference tissue approach. Hence, they emphasize the importance of a per tracer evaluation of suitable RRs.

Conclusion

Outcome measures of all reference regions correlated well with the gold standard and showed stable test-retest performance. However, the largest bias compared with the gold standard was observed for eroded subcortical white matter, followed by whole brain stem and white matter brainstem/pons. Furthermore, using the 60-90 min acquisition window, significant longitudinal alterations in SUV were observed, for whole brain stem and white matter brainstem/pons reference regions. Therefore, grey matter cerebellum and whole cerebellum are considered to be the best RRs for measuring amyloid burden with [11C]PiB.

Acknowledgements

The authors would like to thank the staff of the department of Radiology and Nuclear Medicine of the Amsterdam UMC, location VUmc for skilful acquisition of the scans and Mette Stam for assistance with the plasma input data analyses.

References

1. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol*. 2010;9:119-128.
2. Mallik A, Drzezga A, Minoshima S. Clinical Amyloid Imaging. *Semin Nucl Med*. 2017;47:31-43.
3. Mathis CA, Wang Y, Holt DP, Huang G-F, Debnath ML, Klunk WE. Synthesis and Evaluation of 11C-Labeled 6-Substituted 2-Arylbenzothiazoles as Amyloid Imaging Agents. *J Med Chem*. 2003;46:2740-2754.
4. Rabinovici GD, Jagust WJ. Amyloid Imaging in Aging and Dementia: Testing the Amyloid Hypothesis *In Vivo*. *Behavioural Neurology*. 2009;21:117-128.
5. Berckel BNM van, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. *J Nucl Med*. 2013;54:1570-1576.
6. Yaqub M, Tolboom N, Boellaard R, et al. Simplified parametric methods for [11C]PiB studies. *Neuroimage*. 2008;42:76-86.
7. Carson RE, Channing MA, Blasberg RG, et al. Comparison of Bolus and Infusion Methods for Receptor Quantitation: Application to [18 F]Cyclofoxy and Positron Emission Tomography. *Journal of Cerebral Blood Flow & Metabolism*. 1993;13:24-42.
8. Heeman F, Yaqub M, Alves IL, et al. Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies: *Journal of Cerebral Blood Flow & Metabolism*. April 2020.
9. Gunn RN, Gunn SR, Cunningham VJ. Positron Emission Tomography Compartmental Models. *Journal of Cerebral Blood Flow & Metabolism*. 2001;21:635-652.
10. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. *Neuroimage*. 1996;4:153-158.
11. Cunningham VJ, Hume SP, Price GR, Ahier RG, Cremer JE, Jones AKP. Compartmental Analysis of Diprenorphine Binding to Opiate Receptors in the Rat in vivo and its Comparison with Equilibrium Data in vitro. *Journal of Cerebral Blood Flow & Metabolism*. 1991;11:1-9.
12. Price JC, Klunk WE, Lopresti BJ, et al. Kinetic Modeling of Amyloid Binding in Humans using PET Imaging and Pittsburgh Compound-B. *J Cereb Blood Flow Metab*. 2005;25:1528-1547.
13. Thal DR, Rüb U, Orantes M, Braak H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology*. 2002;58:1791-1800.
14. Wegiel J, Wisniewski HM, Dziewiatkowski J, et al. Cerebellar atrophy in Alzheimer's disease—clinicopathological correlations. *Brain Research*. 1999;818:41-50.
15. Edison P, Hinz R, Ramlackhansingh A, et al. Can target-to-pons ratio be used as a reliable method for the analysis of [11C]PiB brain scans? *NeuroImage*. 2012;60:1716-1723.

16. Bullich S, Villemagne VL, Catafau AM, et al. Optimal Reference Region to Measure Longitudinal Amyloid-β Change with 18F-Florbetaben PET. *J Nucl Med*. 2017;58:1300-1306.
17. Landau SM, Fero A, Baker SL, et al. Measurement of Longitudinal β-Amyloid Change with 18F-Florbetapir PET and Standardized Uptake Value Ratios. *J Nucl Med*. 2015;56:567-574.
18. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia*. 2015;11:1-15.e4.
19. Lowe VJ, Lundt ES, Senjem ML, et al. White Matter Reference Region in PET Studies of 11C-Pittsburgh Compound B Uptake: Effects of Age and Amyloid-β Deposition. *J Nucl Med*. 2018;59:1583-1589.
20. Ossenkoppele R, Tolboom N, Foster-Dingley JC, et al. Longitudinal imaging of Alzheimer pathology using [11C]PiB, [18F]FDDNP and [18F]FDG PET. *Eur J Nucl Med Mol Imaging*. 2012;39:990-1000.
21. Tolboom N, Yaqub M, Boellaard R, et al. Test-retest variability of quantitative [11C]PiB studies in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2009;36:1629-1638.
22. Ossenkoppele R, Prins ND, Pijnenburg YAL, et al. Impact of molecular imaging on the diagnostic process in a memory clinic. *Alzheimers Dement*. 2013;9:414-421.
23. Rask T, Dyrby T, Comerci M, et al. PVElab: Software for correction of functional images for partial volume errors. *Neuroimage*. 2004;22.
24. Hammers A, Allom R, Koepp MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp*. 2003;19:224-247.
25. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. *J Cereb Blood Flow Metab*. 1996;16:834-840.
26. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 2013;49:764-766.
27. Martin Bland J, Altman DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet*. 1986;327:307-310.
28. Jack CR, Wiste HJ, Lesnick TG, et al. Brain b-amyloid load approaches a plateau. 2013:7.
29. Page SW, Maddison JE. Chapter 1 - Principles of clinical pharmacology. In: Maddison JE, Page SW, Church DB, Maddison JE, Page SW, Church DB, eds. Small Animal Clinical Pharmacology (Second Edition). Edinburgh; 2008:1-26.
30. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19:716-723.
31. Hodges JL, Lehmann EL. Estimates of Location Based on Rank Tests. *Ann Math Statist*. 1963;34:598-611.

32. Su Y, Blazey TM, Owen CJ, et al. Quantitative Amyloid Imaging in Autosomal Dominant Alzheimer's Disease: Results from the DIAN Study Group. Herholz K, ed. *PLoS ONE*. 2016;11:e0152082.
33. Schwarz CG, Senjem ML, Gunter JL, et al. Optimizing PiB-PET SUVR Change-Over-Time Measurement by a large-scale analysis of Longitudinal Reliability, Plausibility, Separability, and Correlation with MMSE. *Neuroimage*. 2017;144:113-127.
34. Tryputsen V, DiBernardo A, Samtani M, et al. Optimizing Regions-of-Interest Composites for Capturing Treatment Effects on Brain Amyloid in Clinical Trials. *Journal of Alzheimer's Disease*. 2015;43:809-821.
35. Oliveira F, Leuzy A, Castelhano J, et al. Data driven diagnostic classification in Alzheimer's disease based on different reference regions for normalization of PiB-PET images and correlation with CSF concentrations of Aβ species. *NeuroImage: Clinical*. 2018;20:603-610.
36. Yun HJ, Moon SH, Kim HJ, et al. Centiloid method evaluation for amyloid PET of subcortical vascular dementia. *Scientific Reports*. 2017;7:16322.
37. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet Neurology*. 2013;12:357-367.
38. Jack CR, Wiste HJ, Lesnick TG, et al. Brain b-amyloid load approaches a plateau. 2013:7.
39. Leal SL, Lockhart SN, Maass A, Bell RK, Jagust WJ. Subthreshold Amyloid Predicts Tau Deposition in Aging. *J Neurosci*. 2018;38:4482-4489.
40. Villemagne VL, Bourgeat P, Doré V, et al. Amyloid imaging in therapeutic trials: The quest for the optimal reference region. *Alzheimer's & Dementia*. 2015;11:P21-P22.
41. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. Validation of the Semiquantitative Static SUVR Method for 18F-AV45 PET by Pharmacokinetic Modeling with an Arterial Input Function. *J Nucl Med*. 2017;58:1483-1489.

Supplementary materials

Validation parameter fit boundaries

All data were fitted using SRTM with parameter fit boundaries based on theoretical assumptions, considering grey and white matter reference regions were used. Lower boundaries were the following: R_1: 10^{-7}, k_2: 10^{-7}, BP_{ND}: -1, and upper boundaries: R_1: 10^2, k_2: 10^2, BP_{ND}: 10^2. To optimise the lower and upper boundaries, the minimum and maximum parameter values of a clean dataset were used. This dataset was created by removing all regional fit parameters with a standard deviation (SD) >50%, i.e. if BP_{ND} SD >50%, also the corresponding R_1 and k_2 values were removed (see Supplementary Table 1 below).

Supplementary Table 1. Parameter fit boundaries optimised per reference region

	R_1	k_2	BP_{ND}
GM Cerebellum	0.500	0.011	-0.200
	1.100	0.400	1.600
Whole Cerebellum	0.550	0.014	-0.200
	1.200	0.900	1.350
WM Brainstem/ Pons	0.700	0.011	-0.500
	1.550	2.200	1.100
Whole Brainstem	0.700	0.009	-0.450
	1.800	3.800	1.250
Subcortical Eroded	0.800	0.011	-0.500
WM	2.150	1.350	0.400

Regional outliers

Outliers were observed for the following regions: fusiform gyrus (2), anterior (13) and posterior cingulate gyrus (5), gyrus rectus (3), inferior frontal gyrus (2), inferior lateral remainder of parietal lobe (1), middle and inferior temporal gyrus (1), middle frontal gyrus (3), orbitofrontal gyrus (2), superior frontal gyrus (2), superior parietal gyrus (1), and superior temporal gyrus (1).

Supplementary Table 2a. Bland-Altman: agreement with the gold standard for all reference tissue methods with varying RRs

		DVR_{RLOGAN}	DVR_{SRTM}	$SUVr_{40-60}$	$SUVr_{60-90}$
GM Cerebellum	Bias	4.7	3.3	-0.5	-2.2
	95% CI	[-22.7 32.0]	[-27.1 33.7]	[-34.0 33.0]	[-29.1 24.6]
Whole Cerebellum	Bias	12.4	13.7	10.3	7.7
	95% CI	[-18.1 43.0]	[-19.9 47.3]	[-26.0 46.6]	[-24.4 39.8]
WM Brainstem/ Pons	Bias	39.7	38.3	48.8	39.6
	95% CI	[6.6 72.7]	[2.2 74.3]	[12.3 85.3]	[3.0 76.2]
Whole Brainstem	Bias	36.3	33.3	45.0	36.2
	95% CI	[3.3 69.2]	[-7.0 73.6]	[8.7 81.3]	[0.0 72.3]
Subcortical Eroded WM	Bias	45.4	45.8	53.7	54.7
	95% CI	[15.3 75.4]	[17.5 74.1]	[23.3 84.1]	[27.7 81.8]

Bias: mean % bias as compared with the gold standard, DVR_{PI_GMCB} and 95% confidence interval (CI) calculated as follows: ((gold standard – RT method)/average*100). Data used: 18 target tissue regions derived from six subjects with plasma input data available.

Supplementary Table 2b. Relationship between bias (%) and underlying amyloid burden

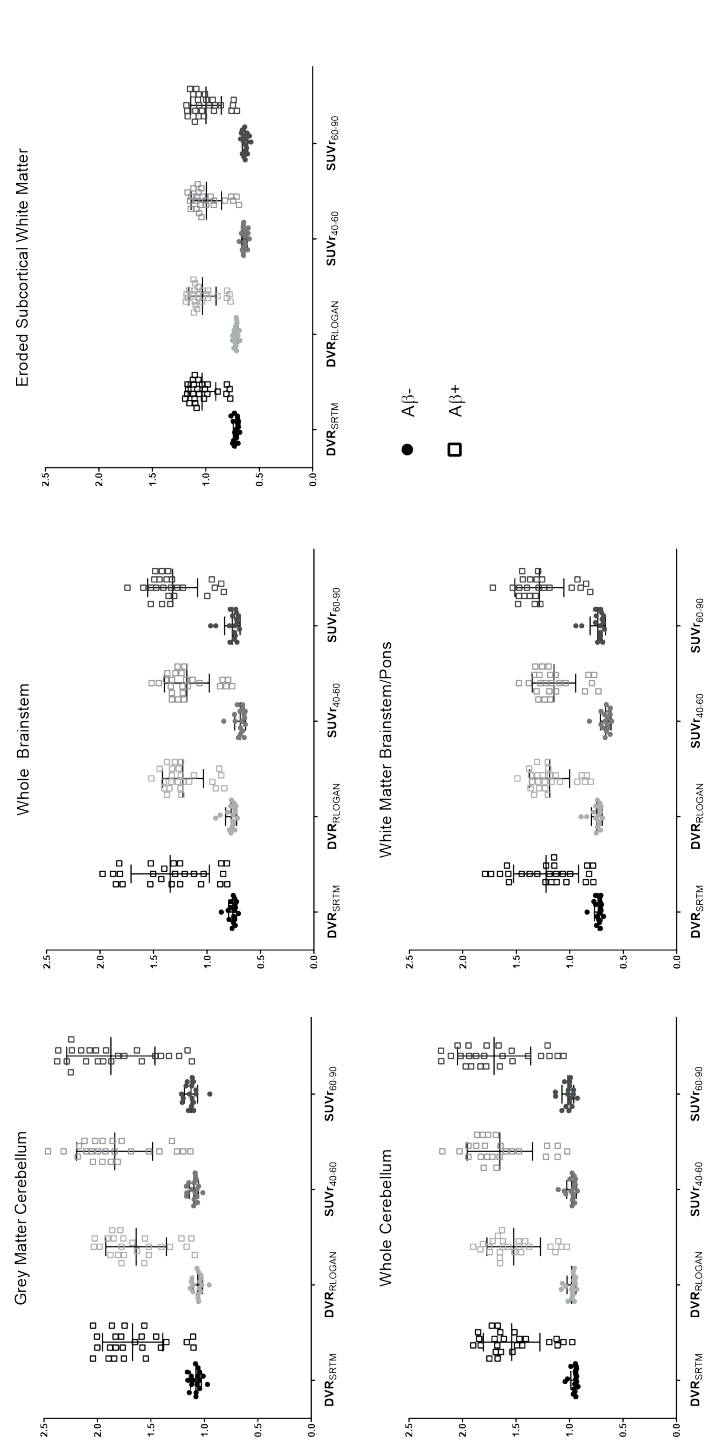
	DVR_{RLOGAN}	DVR_{SRTM}	$SUVr_{40-60}$	$SUVr_{60-90}$
GM Cerebellum	15.6#	12.5#	9.2*	3.4
Whole Cerebellum	16.1#	11.3*	7.8	3.0
WM Brainstem/Pons	11.7*	-22.6#	0.1	-3.3
Whole Brainstem	11.0*	-14.7#	-0.2	-4.3
Subcortical Eroded WM	25.4#	22.8#	13.2*	9.9*

Values correspond to slopes of the linear regression analysis through Bland-Altman data points. Data used: 18 target tissue regions derived from six subjects with plasma input data available.
* $p<0.05$, # $p<0.01$

Supplementary Table 3. Hodges-Lehmann estimate of the median difference between Aβ+ and Aβ- scans

		DVR_{RLOGAN}	DVR_{SRTM}	$SUVr_{40-60}$	$SUVr_{60-90}$
GM Cerebellum	Estimate	-0.7	-0.7	-0.8	-0.9
	LB UB	[-0.7 -0.6]	[-0.7 -0.6]	[-0.9 -0.8]	[-1.0 -0.7]
Whole Cerebellum	Estimate	-0.6	-0.7	-0.8	-0.8
	LB UB	[-0.7 -0.6]	[-0.7 -0.6]	[-0.8 -0.7]	[-0.9 -0.7]
WM Brainstem/Pons	Estimate	-0.5	-0.5	-0.6	-0.6
	LB UB	[-0.5 -0.5]	[-0.7 -0.4]	[-0.6 -0.5]	[-0.7 -0.6]
Whole Brainstem	Estimate	-0.5	-0.6	-0.6	-0.6
	LB UB	[-0.6 -0.5]	[-0.8 -0.5]	[-0.6 -0.5]	[-0.7 -0.6]
Subcortical Eroded WM	Estimate	-0.4	-0.4	-0.4	-0.4
	LB UB	[-0.4 -0.3]	[-0.4 -0.3]	[-0.4 -0.4]	[-0.4 -0.4]

Hodges-Lehmann estimate and its lower and upper boundaries (LB, UB). As null is not within the boundaries, the null-hypothesis of no differences between Aβ+ and Aβ- scans is rejected. All global cortical values belong to baseline scans from the test-retest and longitudinal study (*N*=43)



Supplementary Figure 1. Discriminating between Aβ+ and Aβ- scans Individual global cortical averages were used. Plots show the mean and standard deviation per RR for each of the methods.

PART III

Impact on clinical practise and trials

10.

Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging

Isadora Lopes Alves, **Fiona Heeman**, Lyduine E. Collij, Gemma Salvadó, Nelleke Tolboom, Natàlia Vilor-Tejedor, Pawel Markiewicz, Maqsood Yaqub, David Cash, Elizabeth C. Mormino, Philip S. Insel, Ronald Boellaard, Bart N.M. van Berckel, Adriaan A. Lammertsma, Frederik Barkhof, Juan Domingo Gispert

***As published in Alzheimer's Research & Therapy*, 2021; 13:82**

Abstract

Background: Detecting subtle-to-moderate biomarker changes such as those in amyloid PET imaging becomes increasingly relevant in the context of primary and secondary prevention of Alzheimer's disease (AD). This work aimed to determine if and when distribution volume ratio (DVR; derived from dynamic imaging) and regional quantitative values could improve statistical power in AD prevention trials.

Methods: Baseline and annualized % change in [^{11}C]PiB SUVR and DVR were computed for a global (Cortical) and regional (Early) Composite from scans of 237 cognitively unimpaired subjects from the OASIS-3 database (www.oasis-brains.org). Bland-Altman and correlation analyses were used to assess the relationship between SUVR and DVR. General Linear Models and Linear Mixed Models were used to determine effects of age, sex and *APOE*-ε4 carriership on baseline and longitudinal amyloid burden. Finally, differences in statistical power of SUVR and DVR (Cortical or Early Composite) were assessed considering three anti-amyloid trial scenarios: secondary prevention trials including subjects with 1) intermediate-to-high (Centiloid>20.1), or 2) intermediate (20.1<Centiloid≤49.4) amyloid burden, and 3) a primary prevention trial focusing on subjects with low amyloid burden (Centiloid≤20.1). Trial scenarios were set to detect 20% reduction in accumulation rates across the whole population and in *APOE*-ε4 carriers only.

Results: Although highly correlated to DVR (ρ=0.96), cortical SUVR overestimated DVR cross-sectionally and in annual % change. In secondary prevention trials, DVR required 143 subjects per arm, compared with 176 for SUVR. Both restricting inclusion to individuals with intermediate amyloid burden levels or to *APOE*-ε4 carriers alone further reduced sample sizes. For primary prevention, SUVR required less subjects per arm (N=855) compared to DVR (N =1508) and the Early Composite also provided considerable sample size reductions (N =855 to N=509 for SUVR, N =1508 to N =734 for DVR).

Conclusion: Sample sizes in AD secondary prevention trials can be reduced by the acquisition of dynamic PET scans and/or by restricting inclusion to subjects with intermediate amyloid burden or to *APOE*-ε4 carriers only. Using a targeted Early Composite only leads to reductions of sample size requirements in primary prevention trials. These findings support strategies to enable smaller Proof-of-Concept Phase II clinical trials to better streamline drug development.

Background

With the recently established biological definition of Alzheimer's disease (AD) (*1*) and the increased availability of (imaging) biomarkers, the research community is now well-equipped to study this disease from its earliest pathological changes to later-stage clinical presentations of cognitive impairment (*2*). Especially in the context of much needed treatment and prevention strategies, this research framework can be extremely valuable in accurately identifying individuals in the AD *continuum*, who might benefit from disease-modifying therapies.

With varying degrees of pathological confirmation, recent years have seen many disease-modifying therapies that failed to meet primary endpoints and impact cognitive functioning (*3*). In fact, despite promising signals observed in a number of anti-amyloid clinical trials (*4-8*), the lack of downstream effects on cognition posed important questions on the validity of the widely accepted amyloid cascade hypothesis and highlighted our (still) limited understanding of the mechanisms involved in this disease. Nonetheless, recent results such as those from the aducanumab (*7,9,10*) or BAN2410(*11*) trials have shown promising signals for anti-amyloid therapies, and in fact have encouraged the development of earlier preventive Phase 3 trials focusing on subjects with preclinical AD such as the AHEAD 3-45 Study (*12*). As a result, this shift to prevention in earlier stages of the disease and the (possible) future need for pathological confirmation pre-treatment may increase the use of biomarkers such as amyloid Positron Emission Tomography (PET) imaging for both screening and measurement of treatment effects. However, a marked discrepancy in duration between most short-term studies and the long-term pathological processes such as Aβ plaque accumulation (*13,14*) may result in the need to detect subtle-to-moderate biomarker changes (*15*).

When focusing on the early stages of AD with amyloid PET, observed changes in Aβ burden are mostly focal (*16-18*) and it may be difficult to detect these changes with sufficient statistical power, challenging standard analytical approaches and the traditional use of a global measure of amyloid burden (*19,20*). Recent work suggests that regional amyloid PET assessments can improve early detection of pathology (*16,21,22*), and achieve increased power in clinical trials (*23*). In addition, several PET studies have investigated potential methodological improvements to increase statistical power in longitudinal settings and better discriminate sub-populations cross-sectionally (*24*). These studies generally focused on improving technical factors affecting image quality

such as partial volume effects (*25*) or on modeling and pre-processing choices impacting measurement stability, such as the choice of reference region (*26-28*). However, since the vast majority of PET studies performs static acquisitions, these improvements remain mostly limited to the use of the standard uptake value ratio (SUVR) metric. Although easily available from short static scans, SUVR is a semi-quantitative and biased proxy of the specific amyloid burden as measured by binding potential (BP_{ND}) or distribution volume ratio (DVR) (*29,30*), which are available only from dynamic scans. Specifically, SUVR is known to suffer from technical and physiological sources of bias such as inconsistent scanning window and changes in cerebral blood flow (*30,31*). However, traditional dynamic acquisitions can significantly increase the duration and cost of studies; therefore, compromises have been proposed, such as the collection of early frames in addition to the standard late-uptake image acquisition (*32*). In fact, this early frame collection not only allows for the determination of DVR, but provides an additional parameter (R_1) that can serve as a proxy for cerebral blood flow, another important marker of disease in AD (*33,34*).

Considering current and future research needs, this study aims to determine if and when dynamic imaging and targeted regional quantification could improve statistical power in primary and secondary prevention trials using longitudinal amyloid PET imaging. For that purpose, we estimated the number of participants per arm needed in three hypothetical trial scenarios aiming to reduce amyloid accumulation rates by at least 20%: 1) one in subjects with low amyloid burden for primary prevention, and two for secondary prevention, either 2) including all subjects with abnormal amyloid levels (intermediate-to-high) or 3) focusing on those at the earliest stages of pathology (intermediate levels). We compared the sample sizes required when using SUVR and DVR as amyloid load metric in both the whole population as well in trial scenarios only recruiting *APOE*-ε4 carriers.

Methods

Datasets

This work included two separate datasets: the first was used for main analyses, and the second for calculating test-retest variability for [^{11}C]PiB SUVR and DVR.

For the first dataset, tabulated PET data were obtained from the Open Access Series of Imaging Studies (OASIS-3) dataset, which is a longitudinal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's Disease (www.oasis-brains.org). This dataset is a retrospective compilation of data collected across several ongoing projects through the Washington University of Saint Louis Knight Alzheimer's Disease Research Center (ADRC) over the course of 30 years (*35*). A total of 237 subjects were selected based on 1) being classified as cognitively unimpaired (CU) and 2) having at least two dynamic [^{11}C]PiB PET scans with a minimum of one year between sessions available.

For the second dataset, eleven subjects (4 cognitively unimpaired, 1 mild cognitive impaired, and 6 with AD dementia), were selected from a previously reported test-retest (TRT) study at the Amsterdam University Medical Center location VUmc (*36*). Test and retest scans were performed within a one week interval.

Image acquisition and processing

A brief description of data collection and standard imaging processing pipelines for each dataset can be found below.

OASIS-3 60 minutes dynamic [^{11}C]PiB PET imaged were acquired starting at the intravenous administration of approximately 12mCi of radiotracer. Data was collected in 3D mode on a Siemens/CTI EXACT HR+ scanner or a Biograph 40 PET/CT scanner. Accompanying anatomical T1-weighted MPRAGE MR scans were acquired using either a Siemens 1.5 of 3T scanner. Image processing was performed with a local processing pipeline (PUP; https://github.com/ysu001/PUP), described in detail previously (*37*). In short, the standard FreeSurfer (v5.3; Martinos Center for Biomedical Imaging, Charlestown, Massachusetts, USA; https://surfer.nmr.mgh.harvard.edu/fswiki) based PUP processing includes a scanner resolution harmonization filter (*38*), inter-frame motion correction, PET-MR registration, and regional time-activity curves extraction for all regions from the Desikan-Killiany atlas (DK) (*39*). Using the cerebellar cortex as the reference region, reference Logan graphical analysis (RLogan) (*40*) was used to determine DVR with t^* set to 30 min post-injection (p.i.). In parallel, SUVR was extracted for the same time-window of 30-60 min p.i.

For the TRT study, 90 minutes dynamic [^{11}C]PiB PET scans were performed on a Siemens ECAT EXACT HR+ scanner and a structural T1-weighted MR scan on a 1.5T Siemens

Sonata scanner. First, structural T1-weighted MR images were co-registered to the PET scan using Vinci software (Max Planck Institute for Neurological Research, Cologne, Germany) and PVE-lab software was used to extract the cerebellar cortex time-activity curve based on the Hammers atlas (*41,42*). Next, both DVR (RLogan) and SUVR were calculated from 30-60 min p.i. in order to compare results with those from the OASIS-3 dataset, and finally normalized to the cerebellar cortex using PPET software (*43*). These parametric images were then warped into MNI space using SPM12 and the DK atlas was used to extract regional SUVR and DVR values.

Both global and regional analyses were performed on the SUVR and DVR data. A global measure of amyloid burden was determined based on a "Cortical Composite" created from grey-matter FreeSurfer-defined frontal, parietal, temporal and precuneus regions (37). In addition, an "Early Composite" was defined from three grey-matter DK regions, namely the isthmus cingulate, precuneus and lateral orbitofrontal cortices. These regions were chosen based on literature for consistently displaying increased amyloid burden in early disease stages, as well as higher rates of accumulation compared to cortical composites (*16-19,44*). Finally, corresponding and previously validated Centiloid (CL) values were also available for comparison in the OASIS-3 dataset (*26*).

Levels of β-amyloid burden

Three different levels of amyloid burden were defined based on CL cut-offs available from literature and validated against pathology (*45*). Low amyloid burden was defined as CL values ≤ 20.1, a threshold showing the highest accuracy in detecting moderate or frequent plaque density. In contrast, high amyloid burden was defined as CL values above 49.4, the threshold found to identify intermediate or high likelihood of Alzheimer's disease according to NIA-AA 2012 criteria (*46*). Finally, intermediate levels were those with $20.1 < CL \leq 49.4$.

Amyloid accumulation

In order to account for differences in number of scans and interval between visits, a linear mixed effects model (LME) with random intercepts and random slopes was used to determine annualized rates of Aβ accumulation for every metric (SUVR and DVR) in the OASIS-3 dataset. To facilitate interpretability when reporting results, these were also normalized to baseline Aβ levels and will be reported as annualized % change.

Next, the TRT variability of each quantitative metric derived from the TRT dataset was used as a cut-off to determine the proportion of subjects to be considered as "accumulators", i.e. those with annualized % change above TRT variability. Relative TRT variability was calculated for all subjects from the TRT dataset (N=11) and for cognitively unimpaired subjects only (N=4), according to Equation 1, where the estimate of amyloid burden (DVR or SUVR) of the test scan is denoted as T and for the retest scan as R.

$$TrT\ variability\ (\%) = \frac{|T - R|}{0.5 \cdot |T + R|} \cdot 100 \qquad (1)$$

Statistical analysis

All statistical analysis were performed using R Statistical Software (version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria). Results are reported as mean ± standard deviation (μ ± SD) or median (M) and interquartile range (IQR), as appropriate. In all analyses, DVR was considered the reference metric.

To assess the relationship between cortical SUVR and DVR at baseline and longitudinally, Bland-Altman plots, correlation analyses, and paired t-tests (or Wilcoxon Signed-Rank Test) were used. In addition, paired t-tests (or Wilcoxon Signed-Rank Test) were also used to assess differences between a Cortical Composite and an Early Composite in the estimation of amyloid burden and accumulation rates. To assess the relationship between baseline amyloid burden and longitudinal amyloid accumulation, a linear, a quadratic and a natural cubic spline model with one knot were tested, and the optimal model was determined based on the Akaike Information Criteria (AIC). Finally, effects of age, *APOE*-ε4 carriership (presence of at least one ε4 allele), and sex on baseline amyloid burden were assessed by a General Linear Model (GLM). Similarly, a Linear Mixed Effects Model (LME) was used to determine the effect of the same variables on amyloid accumulation, accounting for baseline amyloid burden.
The analyses above were performed in order to determine the generalizability of the OASIS-3 dataset with respect to other cohorts, such that the results of the sample size calculations can be contextualized appropriately.

Sample size calculations

Using the LME estimates for annualized accumulation rates and respective standard deviations, the *sampsizepwr* function in Matlab (1-β = 80% power and a two-tailed t-test type-I error of α=0.05) was used to determine sample sizes required to detect differences in accumulation rates in three hypothetical 12-month placebo-controlled randomized

anti-amyloid clinical trials. The trial designs assumed participants undergo a PET scan at baseline and another at the completion of the trial. These were computed separately for SUVR and DVR, using the Cortical Composite and the Early Composite, both across the whole population and restricted to *APOE*-ε4 carriers only.

The tested trial scenarios were the following:

1. A secondary prevention trial aiming to detect a 20% reduction in β-amyloid accumulation rates in individuals with intermediate-to-high amyloid burden (CL > 20.1) at baseline;
2. An *earlier* secondary prevention trial aiming to detect a 20% reduction in β-amyloid accumulation rates focusing in individuals with intermediate amyloid burden (20.1 < CL ≤ 49.4) at baseline;
3. A primary prevention trial aiming to detect a 20% reduction in in β-amyloid accumulation rates in individuals with low amyloid burden (CL ≤ 20.1) at baseline.

Results

On average, OASIS-3 subjects underwent 2.5 ± 0.6 scans [range 2-5], with an average of 4.8 ± 2.1 years between the first and the last scan [range 1-9.6]. The majority of subjects were female (65.0%), 32.9% of them were *APOE*-ε4 carriers, and the mean age at the time of the first PET session was 65.3 ± 9.4 years. Complete OASIS-3 cohort demographics are shown in Table 1.

Similarly, the cognitively unimpaired individuals from the TRT dataset (N=4) were mainly female (75.0%), 33.3% were *APOE*-ε4 carriers, and their mean age was 66.8 ± 4.1 years. In contrast, the full dataset (N=11) had a higher proportion of *APOE*-ε4 carriers (62.5%), equivalent proportion of males and females (45.5% females), and a mean age of 64.0 ± 4.9 years. Of note, *APOE* genotyping was missing for 3/11 subjects.

Cortical and regional β-amyloid quantification

Baseline cortical SUVR (M=1.08, IQR=1.14-1.05) and DVR (M=1.05, IQR=1.10-1.02) were highly correlated (ρ=0.96), and SUVR consistently overestimated DVR (Figure 1A). Assuming DVR as the reference standard, this bias in SUVR was proportional to the underlying level of amyloid burden, where an increase in one amyloid burden unit translated to a 30% increase in bias (slope in Figure 1B). The same pattern was found for

annual % change, where the two metrics were highly correlated (r=0.98, Figure 1C), but SUVR (μ= 1.06 ± 1.30%) overestimated DVR-based accumulation rates (μ= 0.75 ± 1.11%) by 15% at every unit increase in underlying accumulation rates (slope in Figure 1D).

The relationship between baseline and longitudinal cortical amyloid burden was well described by a quadratic model for both SUVR (R^2=0.21, ΔAIC_{linear}=-94.4, ΔAIC_{spline}=-95.3) and DVR (R^2=0.26, ΔAIC_{linear}=-78.8, ΔAIC_{spline}=-79.2) (Figure 2A), where subjects in the intermediate amyloid burden group displayed the highest accumulation rates on average (Table 1). Across the whole cohort, baseline SUVR and DVR and respective annual % change did not differ between males and females, while higher age was associated with higher baseline SUVR (β=0.007, t=4.60, p<0.001) and DVR (β=0.005, t=4.17, p<0.001), but did not predict accumulation rates. Similarly, *APOE*-ε4 carriership was associated with higher baseline levels of amyloid burden (SUVR: β=0.163, t=5.69, p<0.001; DVR: β=0.122, t=5.53, p<0.001), and only using SUVR was it also related to higher accumulation rates (β=0.014, t=2.55, p=0.011).

As expected, both baseline and accumulation rates with SUVR and DVR were significantly higher when using the Early Composite compared to the Cortical Composite (Table 1).

Table 1. Descriptive composition of included subjects in terms of baseline amyloid status, *APOE*-ε4 carriership, sex, age and annualized accumulation rates in DVR and SUVR.

	Whole cohort			Low burden group (CL ≤ 20.1)			Intermediate burden group (20.1 < CL ≤ 49.4)			High burden group (CL > 49.4)		
	All	*APOE*-ε4 non-carriers	*APOE*-ε4 carriers	All	*APOE*-ε4 non-carriers	*APOE*-ε4 carriers	All	*APOE*-ε4 non-carriers	*APOE*-ε4 carriers	All	*APOE*-ε4 non-carriers	*APOE*-ε4 carriers
Number of subjects (%)	237 (100.0)	159 (67.1)	78 (32.9)	194 (81.8)	146 (75.2)	48 (24.8)	20 (8.4)	5 (25.0)	15 (75.0)	23 (9.7)	8 (34.8)	15 (65.2)
Number of women (%)	154 (65.0)	100 (62.9)	54 (69.2)	129 (66.5)	95 (65.1)	34 (70.8)	13 (65.0)	1 (20.0)	12 (80.0)	12 (52.2)	4 (50.0)	8 (53.3)
Age, y	65.3 ± 9.4	65.8 ± 9.5	64.5 ± 9.2	64.4 ± 9.6	65.3 ± 9.7	61.7 ± 9.2	67.6 ± 7.4	67.7 ± 3.9	67.6 ± 8.3	71.1 ± 6.6	72.4 ± 7.0	70.3 ± 6.5
SUVR (Cortical Composite)												
Baseline (IQR)	1.08 (1.05-1.14)	1.07 (1.05-1.12)	1.15 (1.05-1.44)	1.07 (1.04-1.11)	1.07 (1.04-1.10)	1.07 (1.04-1.14)	1.40 (1.29-1.45)	1.38 (1.30-1.47)	1.41 (1.28-1.45)	1.73 (1.61-1.91)	1.79 (1.59-1.91)	1.72 (1.61-1.91)
Annual % change	1.06 (1.30)	0.77 (1.08)	1.63 (1.53)	0.80 (1.18)	0.69 (0.99)	1.15 (1.56)	2.83 (0.93)	2.81 (0.69)	2.84 (1.02)	1.77 (1.18)	2.37 (1.34)	1.98 (1.07)
Effect size	0.76	0.69	1.10	0.69	0.63	0.72	3.54	4.75	3.00	1.58	1.05	1.89
Number of accumulators (%) [CUs TRT]	45 (23.6)	25 (15.7)	31 (39.7)	27 (13.9)	17 (11.6)	10 (20.8)	17 (85.0)	5 (100.0)	12 (80.0)	12 (52.2)	3 (37.5)	9 (60.0)
Number of accumulators (%) [ALL TRT]	20 (8.4)	5 (3.1)	15 (19.2)	10 (5.2)	3 (2.1)	7 (14.6)	7 (35.0)	1 (20.0)	6 (40.0)	3 (13.0)	1 (12.5)	2 (13.3)
SUVR (Early Composite)												
Baseline (IQR)	1.16 (1.08-1.23)	1.14 (1.11-1.18)	1.22 (1.13-1.54)	1.14 (1.11-1.18)	1.14 (1.11-1.17)	1.15 (1.11-1.20)	1.50 (1.41-1.54)	1.50 (1.42-1.53)	1.50 (1.38-1.55)	1.79 (1.73-1.99)	1.89 (1.78-2.03)	1.79 (1.71-1.98)
Annual % change	1.25 (1.26)	0.99 (1.08)	1.75 (1.44)	1.02 (1.16)	0.91 (1.03)	1.34 (1.44)	2.87 (1.02)	2.74 (0.90)	2.90 (1.08)	1.74 (1.13)	1.36 (1.17)	1.94 (1.09)
Effect size	0.94	0.86	1.20	0.86	0.87	0.94	3.23	3.41	3.00	1.55	1.18	1.84
DVR (Cortical Composite)												
Baseline (IQR)	1.05 (1.02-1.10)	1.04 (1.01-1.07)	1.10 (1.03-1.30)	1.04 (1.01-1.06)	1.04 (1.01-1.06)	1.04 (1.02-1.09)	1.26 (1.21-1.30)	1.29 (1.21-1.34)	1.25 (1.20-1.30)	1.54 (1.43-1.69)	1.59 (1.44-1.71)	1.53 (1.43-1.69)
Annual % change	0.75 (1.11)	0.50 (0.91)	1.25 (1.30)	0.49 (0.96)	0.40 (0.82)	0.75 (1.29)	2.34 (0.72)	2.53 (0.43)	2.34 (0.80)	1.55 (0.93)	1.16 (1.07)	1.76 (0.80)
Effect size	0.69	0.54	1.00	0.48	0.50	0.57	3.75	7.11	3.22	1.86	1.06	2.25
Number of accumulators (%) [CUs TRT]	81 (34.2)	39 (24.5)	42 (53.8)	44 (22.7)	29 (19.9)	15 (31.3)	20 (100.0)	5 (100.0)	15 (100.0)	17 (73.9)	5 (62.5)	12 (80.0)
Number of accumulators (%) [ALL TRT]	33 (13.9)	11 (6.9)	22 (28.2)	13 (6.7)	6 (4.1)	7 (14.6)	13 (65.0)	4 (80.0)	9 (60.0)	7 (30.4)	1 (12.5)	6 (40.0)
DVR (Early Composite)												
Baseline (IQR)	1.11 (1.08-1.17)	1.10 (1.07-1.13)	1.16 (1.09-1.38)	1.09 (1.07-1.12)	1.09 (1.07-1.12)	1.10 (1.07-1.14)	1.34 (1.31-1.39)	1.39 (1.32-1.41)	1.34 (1.27-1.38)	1.66 (1.54-1.77)	1.72 (1.56-1.82)	1.59 (1.51-1.70)
Annual % change	0.94 (1.08)	0.72 (0.94)	1.34 (1.22)	0.71 (0.97)	0.63 (0.87)	0.94 (1.19)	2.49 (0.76)	2.55 (0.71)	2.47 (0.80)	1.50 (0.89)	1.15 (0.96)	1.68 (0.83)
Effect size	0.78	0.72	1.12	0.72	0.70	0.76	3.67	3.89	3.43	1.71	1.19	2.17

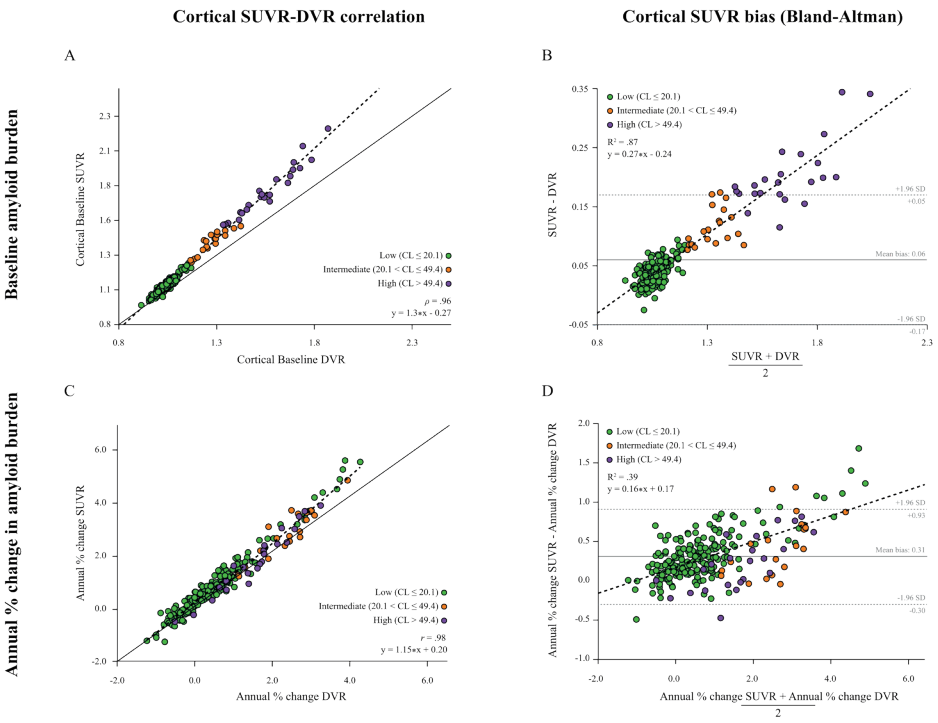


Figure 1. Relationship between SUVR and DVR. On the top panel, a scatterplot between baseline cortical SUVR and DVR across all subjects, with a solid identity line as reference **(a)**, and a Bland-Altman plot displaying a linear relationship between SUVR bias and underlying amyloid burden **(b)**. On the bottom panel, a scatterplot between annualized % cortical SUVR and DVR across all subjects, with a solid identity line as reference **(c)**, and a Bland-Altman plot displaying a linear relationship between bias in annualized % cortical SUVR and underlying accumulation rates, with a dotted line representing a linear regression through the data points **(d).**

TRT and longitudinal amyloid accumulation

In order to assess the proportion of OASIS-3 participants with accumulation rates beyond TRT variability, we determined cut-offs for accumulation based on a separate local TRT dataset.

Focusing on cognitively unimpaired individuals only (*N*=4), SUVR TRT was 1.61% for the Cortical Composite, compared with 0.85% for DVR. Similarly, a TRT of 3.46% was observed for SUVR with an Early Composite, while for the same ROI, DVR TRT was only 2.05%. In addition, when assessing subjects across diagnostic groups (*N*=11), the

pattern remains, with DVR TRT always lower than SUVR (Cortical Composite: 2.12% DVR / 3.45% SUVR, Early Composite: 2.14% DVR / 4.16% SUVR).

Using TRT from cognitively unimpaired subjects as our main cut-off for accumulation (due to cohort comparability) and a Cortical Composite for quantification, 81 (34.2%) individuals were classified as accumulators using DVR compared to 45 (23.6%) using SUVR (Figure 2B). A total of 25 subjects were accumulators with DVR but not SUVR; 17 of them belonging to the low, 3 to the intermediate, and 5 to the high amyloid burden group (Table 1). Similarly, using the Early Composite for quantification and TRT cut-off, SUVR analyses classified 8 (3.4%) of subjects as accumulators compared to 39 (16.5%) when using DVR. In this case, 31 subjects were accumulators with DVR but not with SUVR, 10 of which were from the low, 15 from intermediate, and 6 from the high amyloid burden group.

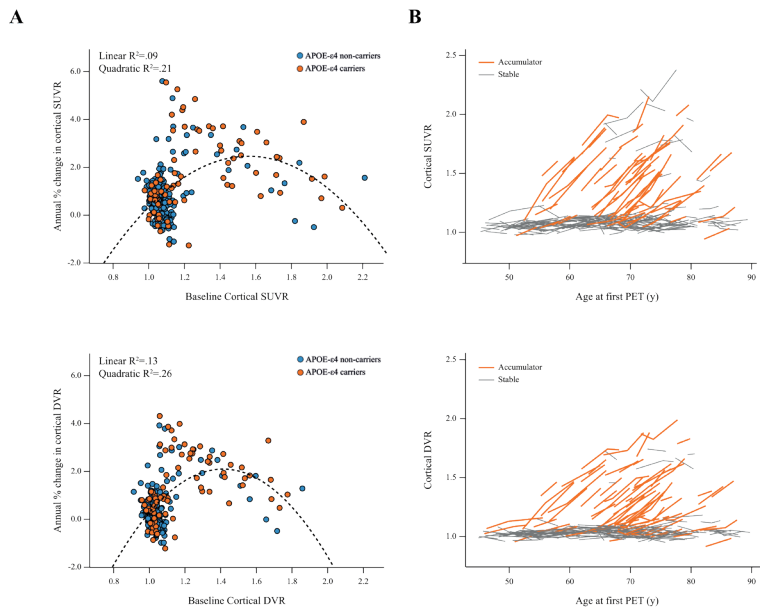


Figure 2. Amyloid accumulation with SUVR and DVR Scatter plot of the relationship between annual % change and baseline amyloid levels using SUVR (top left) and DVR (bottom left) in *APOE*-ε4 carriers (orange) and non-carriers (blue), with a dotted line representing the quadratic model fit **(a)**. Plot of the absolute change in SUVR (top right) and DVR (bottom right) in time, coded for whether subjects were classified as accumulators based on TRT from cognitively unimpaired individuals (orange) or were considered stable (grey) **(b).**

Sample sizes in longitudinal studies

Table 2 summarizes the required sample sizes for three hypothetical trial scenarios, considering different choices with respect to acquisition protocol (static/SUVR or dynamic/DVR), methodological (Cortical Composite or Early Composite), and inclusion criteria (whole population or *APOE*-ε4 carriers only).

For secondary prevention trials aiming to detect a 20% reduction in β-amyloid accumulation rates, the sample sizes required are consistently lower when using DVR compared to SUVR (Table 2), likely because the smaller standard deviation and better TRT observed with DVR outweighs its lower average rate of accumulation (Table 1). In addition, including only *APOE*-ε4 carriers provided considerable reduction in the required sample sizes (whole population: N_{SUVR}=176, N_{DVR}=143, *APOE*-ε4 carriers only: N_{SUVR}=116, N_{DVR}=83), for either region of interest chosen for analysis.

Table 2. Sample size requirements per trial arm, for three hypothetical trial scenarios, comparing differences between using DVR/SUVR, a Cortical/Early Composite ROI, and restricting the inclusion to *APOE*-ε4 carriers or not.

	Whole population				*APOE*- ε4 carriers only			
	SUVR		DVR		SUVR		DVR	
	Cortical ROI	Early ROI	Cortical ROI	Early ROI	Cortical ROI	Early ROI	Cortical ROI	Early ROI
Secondary prevention to detect 20% reduction in accumulation (CL > 20.1)	176	167	143	140	116	125	83	97
Early secondary prevention to detect 20% reduction in accumulation (20.1 < CL ⩽ 49.4)	44	51	39	38	52	56	47	43
Primary prevention to detect 20% reduction in accumulation (CL ⩽ 20.1)	855	509	1508	734	724	455	1162	630

Further, if this secondary prevention trial included only individuals at an earlier stage of the disease (i.e. those with intermediate amyloid burden and thus more likely to have higher accumulation rates), a 4-fold reduction in required sample sizes (N_{SUVR}=44, N_{DVR}=39) can be achieved compared to including subjects from the general population (N_{SUVR}=176, N_{DVR}=143). In both secondary prevention scenarios the use of an Early Composite did not reduce the required sample sizes.

Finally, a primary prevention trial required the largest sample sizes overall as expected, and the use of an Early Composite reduced the number of subjects needed to detect the desired effect by ~40-50%, in case of both SUVR ($N_{CORTICAL}$=855, N_{EARLY}=509) and DVR ($N_{CORTICAL}$=1508, N_{EARLY}=734). Similarly, restricting the trial to *APOE*-ε4 carriers provided approximately ~20% reductions in sample size requirements with either acquisition protocol. However, in this scenario, the use of SUVR provided smaller sample size requirements than DVR (Figure 3), which relates to its higher accumulation rates and similar standard deviation (Table 1).

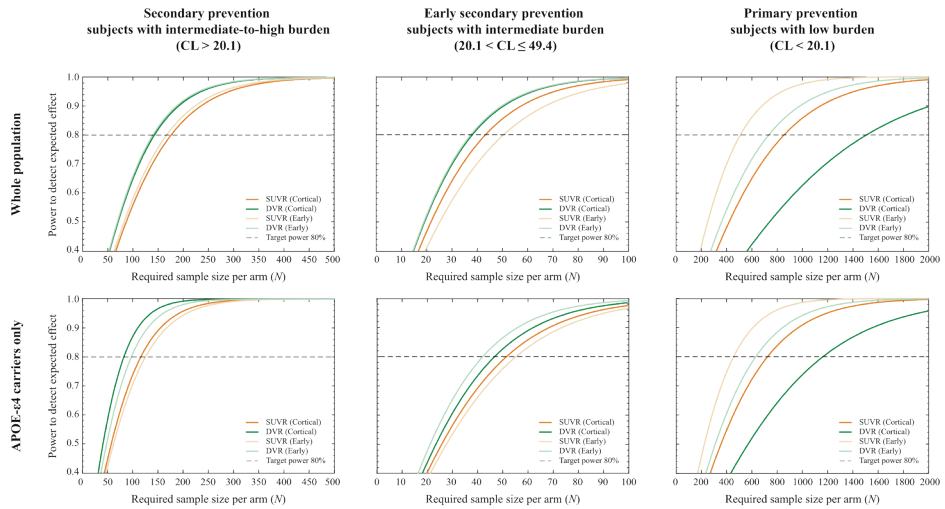


Figure 3. Sample size requirements (per arm)
Relationship between achieved statistical power and number of participants required in three anti-amyloid hypothetical trial scenarios for the general population (top row) or focusing on *APOE*-ε4 carriers only (bottom row). The dotted line represents the desired power of 1-β = 80%.

Discussion

In this work, we observed that the smaller variability of DVR compared to SUVR results in smaller sample size requirements for anti-amyloid secondary prevention trials when using dynamic amyloid PET scans. In addition, focusing on individuals with intermediate levels of amyloid burden who are at the peak of accumulation provides a 4-fold reduction in sample sizes compared to traditional secondary prevention trials (where inclusion criteria includes amyloid-positive individuals regardless of the extent of pathology). As expected, primary prevention trials require larger sample sizes to

achieve similar statistical power, but this can be mitigated by targeting inclusion criteria to *APOE*-ε4 carriers and/or by using an Early Composite region of interest.

First, the direct comparison between dynamic and static parameters in this work confirmed that SUVR largely overestimates DVR, and that this bias is strongly dependent on the underlying levels of amyloid burden (Figures 1A-B). In addition, this overestimation relates to the underlying radiotracer kinetics and can be further influenced by scan time, as well as known confounding effects such as changes in blood flow and tracer clearance *(29,30)*. Especially in the case of disease-modifying therapies, an intervention could affect cerebral blood flow and therefore falsely inflate treatment effects when measured by SUVR (*31*), challenging the interpretation of SUVR-based rates of amyloid accumulation. As a consequence, the results of our primary prevention trial scenario should be interpreted with caution, where the increased accumulation rates observed with SUVR seem to facilitate the detection of treatment effects compared to DVR, despite the increased variability (Tables 1-2). Especially in these early stages of disease where the underlying amyloid PET signal is low, the relatively large contribution of physiological and methodologically-driven fluctuations in the PET signal can lead to misinterpreted results. This is of particular relevance when the tested intervention may impact cerebral blood flow.

In contrast, secondary prevention trials seem to benefit from the acquisition of dynamic scans, where consistent reductions in sample sizes are observed (Table 2). There, the overestimation of SUVR accumulation rates is less pronounced with respect to its increased variability, resulting in a direct improvement in statistical power when using DVR, a metric with overall lower TRT variability (*36*). This finding is in line with a recent publication on tau tracer [^{18}F]flortaucipir, where the differences TRT variability between SUVR and BP_{ND} also led to smaller sample size requirements when using the latter as quantitative metric (*47*). Naturally, obtaining DVR estimates would imply the acquisition of dynamic scans, which can result in a non-negligible increase in patient discomfort, use of scanner time, and overall study cost. To our knowledge, the only available report on the willingness of participants to undergo a second dynamic scan indicates that, at least when using a dual-time window protocol, only 5% of them would consider dropping the study due to discomfort (*48*). Further, considering average rates of €750 for a static scan and €1050 for a dynamic scan (available from the AMYPAD Consortium, data not shown), our results indicate that performing dynamic scans may not significantly impact study costs (DVR: *N*=143, €150k, SUVR: *N*=176, €132k). Therefore, while maintaining

similar cost, the acquisition of dynamic scans can increase statistical power, provide additional biomarker information on cerebral blood flow (*33,34*), and expose less participants to radiation, an ethical consideration that should not be disregarded (*29*).

In addition to the increased statistical power of DVR, focusing subject selection in secondary prevention trials to individuals at the peak of amyloid accumulation ($20.1 < CL \leq 49.4$) provided a 4-fold reduction in required sample sizes (Table 2). In fact, similar results have been reported by Guo and colleagues, who demonstrated that prevention trials must account for the differences in amyloid accumulation phases (Figure 2A) by narrowing the range of amyloid burden in inclusion criteria range; otherwise, estimates of treatment effect can be significantly biased (*49*). Importantly, the interval of amyloid burden used in our work captures the typical range of amyloid positivity cut-offs derived from visual assessment (*45,50-52*), while the upper values around 49.4 CL mostly correspond to levels found in subjects with a clinical presentation of AD (*45,53*). In addition, the range of amyloid burden used in this work for each of the secondary prevention trials are in line to with both the A3 (20-40 CL) and the A45 ($CL > 40$) trials, both of which target a similar population to the OASIS-3 dataset (*11*). Together, our findings further stress the advantages of refining the range of amyloid burden in entry criteria and support the current and future design of smaller, Phase-II, Proof-of-Concept prevention trials in at-risk populations (*54*). Of note, these considerations should be weighed against possible higher screening failure rates.

Interestingly, the secondary prevention trial designs tested in this work did not seem to benefit from the use of an Early Composite ROI. At this stage, the amyloid accumulation in a (global) cortical composite has reached similar rates as those observed in the early regions and has the advantage of larger volume and better count statistics (Table 1). This suggests that, already at the intermediate amyloid burden level, accumulation rates of other regions start to increase and contribute to the global signal. In line with our findings, a previous report described that at higher levels of amyloid burden, the set of regions with increased accumulation rates fall outside of the typical-AD topography (*49*). In contrast, primary prevention trials seem to greatly benefit from the use of an Early Composite ROI, where we observed a ~40-50% reduction in expected sample sizes using a ROI composed of precuneus, isthmus cingulate and lateral orbitofrontal regions (Table 2). These findings are corroborated by a recent report from Insel and colleagues using the Alzheimer's Disease Neuroimaging Initiative dataset (*23*). There, authors showed a reduction of ~62% in required sample sizes when using an early ROI

composed of precuneus and posterior cingulate. Both early regions proposed by Insel's and our work, as well as the late ones described by Guo and colleagues are in excellent agreement with recently proposed amyloid burden staging systems (*16,19,44*). Thus, these findings indicate that in order to significantly impact statistical power, the choice of regions for quantification must be informed by the disease stage of the target population.

Finally, we demonstrated that screening for risk factors such as age and *APOE*-ε4 carriership could further reduce sample size requirements. As expected, age was associated with higher baseline levels of amyloid burden. However, it was not predictive of accumulation rates, which reiterates this is a risk factor for amyloid pathology but does not directly influence the overall accumulation process, as previously suggested in a meta-analysis (*13*). Similarly, *APOE*-ε4 carriership was more frequent in subjects with intermediate-to-high amyloid burden, and carriers were younger than their non-carrier counterparts (Table 1). In addition, carriership was only marginally associated with increased accumulation rates, similar to previous work (*55,56*), an effect which only reached significance for SUVR (likely due to the proportional bias of this metric which increases for higher levels of amyloid and accumulation rates, see Figures 1B and 1D). Together, this suggests *APOE*-ε4 mainly impacts the onset of amyloid pathology rather than the speed of the subsequent accumulation process (*57*). These results are in line with several previous reports, which indicate that even in cognitively unimpaired individuals, *APOE* genotype has a substantial effect on the age-related prevalence of AD pathology (*13,58*). In our work, we find that both primary and secondary prevention trials can still significantly reduce required sample sizes when enrolling *APOE*-ε4 carriers alone, despite their younger age. Therefore, enrichment strategies in a general population could focus on older individuals, while specifically targeting *APOE*-ε4 carriers may allow for the inclusion of younger subjects, as these would already have an increased probability of being in the AD continuum. However, such a strategy may impact both screen failure and future labelling of the drug, restricting its prescription from the general population.

It is important to note that all results in this work relate to a fixed effect (20%) of reducing the accumulation rates in amyloid PET scans, which may seem disconnected from the level of amyloid removal observed in recent anti-amyloid immunotherapies (*7,59*). Indeed, most anti-amyloid trials demonstrate such large differences from baseline in amyloid burden that the effects can even be appreciated visually. Nonetheless, other interventions may have more subtle effects on amyloid burden, either directly or indirectly. Some examples would be BACE1 inhibitors (*60*), drugs with other targets which have downstream amyloid

effects (61), or even non-pharmacological therapies and multi-domain preventive trials such as those being tested in World-Wide FINGERS (*62,63*). As such, 20% reduction of amyloid accumulation may be a relevant target to detect, especially in a short 1-year Proof-of-Concept study. Nonetheless, the overall sample size impacts of using SUVR/DVR, Early/Cortical Composites, or restricting inclusion criteria can still be observed for larger treatment effects (Supplementary Figure 1). Naturally, these differences become less relevant as the expected reductions become larger.

Methodological issues need to be considered when interpreting the findings of this study. First, while DVR is used as the standard of truth in this work, the chosen imaging window for analysis (30-60 min p.i.) and the use of RLogan could both have affected the results of the comparison between SUVR and DVR. Previous studies have indicated that, prior to the 40-50 min interval, [^{11}C]PiB SUV may still be rapidly changing and equilibrium is still not reached. Therefore, this earlier imaging window does not correspond to secular equilibrium conditions, which could have inflated possible flow effects in SUVR and affected RLogan estimates (*64*). In addition, RLogan is known to underestimate true binding potential and suffer from noise-induced bias, while other methods such as SRTM2 and MRTM2 have been proposed as optimal for [^{11}C]PiB and might have produced higher accumulation rates with DVR (*65*). It should also be noted that TRT values from a small single-centre study may not translate to the data collected in OASIS-3. However, the differences between SUVR and DVR TRT reported in this work are in line with previous findings with the same tracer (*66*), as well as with other tracers (*47*). Moreover, the TRT dataset analysed in this work was used as supporting evidence for the superior statistical properties of dynamic PET scans, and the use of literature values would have resulted in equivalent results.

Limitations include the single-tracer character of the study and the relatively limited availability of follow-up data with more than two time points. In addition, one must consider whether the population of OASIS-3 is representative of the primary/secondary prevention trial populations. First, the age range in this work might be too large, but the vast majority of subjects (71%) were between 60 and 85 years of age (*11*). Of note, these results may not be comparable to other tracers, as the kinetics of [^{11}C]PiB are markedly faster than what is observed with e.g. the commercially available F-18 tracers such as [^{18}F]flutemetamol and [^{18}F]florbetaben, which may display even larger biases between SUVR and DVR and therefore also larger differences in sample size requirement between the metrics. This remains to be confirmed and will be explored within the Amyloid

Imaging to Prevent Alzheimer's Disease (AMYPAD) Consortium (*67*). Finally, future work in a larger dataset may consider estimating the uncertainty around sample size estimates to better understand the generalizability of these results and relate them to changes in cognitive functioning, which remains the main outcome measure in most preventive trials to date.

Conclusion

Strategies to improve statistical power differ between secondary and primary AD prevention trials. First, the acquisition of dynamic PET scans can provide reduction in sample sizes only in secondary prevention trials, representing a reasonable alternative to static imaging while reducing the need for exposing healthy participants to ionizing radiation. In contrast, the use of an early composite seems to only benefit primary prevention trials, suggesting that regional analyses must be informed by disease stage in order to provide improved statistical power to trials. Overall, refining inclusion criteria can result in considerable reductions in sample size requirements by identifying individuals at the peak of amyloid accumulation and/or restricting trials to *APOE*-ε4 carriers. These results may provide guidance on how to design smaller Phase II Proof-of-Concept trials without penalizing statistical power to detect treatment-related changes in amyloid accumulation.

Acknowledgements

The authors would like to acknowledge the OASIS-3 project. This research would not have been possible without the time and dedication of every participant included in this study.

References

1. Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology. 2016;87(5):539-47.
2. Jack CR, Jr., Therneau TM, Weigand SD, Wiste HJ, Knopman DS, Vemuri P, et al. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. JAMA Neurol. 2019.
3. Gold M. Phase II clinical trials of anti-amyloid beta antibodies: When is enough, enough? Alzheimers Dement (N Y). 2017;3(3):402-9.
4. Vandenberghe R, Rinne JO, Boada M, Katayama S, Scheltens P, Vellas B, et al. Bapineuzumab for mild to moderate Alzheimer's disease in two global, randomized, phase 3 trials. Alzheimers Res Ther. 2016;8(1):18.
5. Honig LS, Vellas B, Woodward M, Boada M, Bullock R, Borrie M, et al. Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease. N Engl J Med. 2018;378(4):321-30.
6. Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med. 2014;370(4):322-33.
7. Sevigny J, Chiao P, Bussiere T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces Abeta plaques in Alzheimer's disease. Nature. 2016;537(7618):50-6.
8. Egan MF, Kost J, Tariot PN, Aisen PS, Cummings JL, Vellas B, et al. Randomized Trial of Verubecestat for Mild-to-Moderate Alzheimer's Disease. N Engl J Med. 2018;378(18):1691-703.
9. 221AD301 Phase 3 Study of Aducanumab (BIIB037) in Early Alzheimer's Disease [Available from: https://ClinicalTrials.gov/show/NCT02477800.
10. 221AD302 Phase 3 Study of Aducanumab (BIIB037) in Early Alzheimer's Disease [Available from: https://ClinicalTrials.gov/show/NCT02484547.
11. A Study to Evaluate Safety, Tolerability, and Efficacy of BAN2401 in Subjects With Early Alzheimer's Disease [Available from: https://ClinicalTrials.gov/show/NCT01767311.
12. AHEAD 3-45 Study: A Study to Evaluate Efficacy and Safety of Treatment With BAN2401 in Participants With Preclinical Alzheimer's Disease and Elevated Amyloid and Also in Participants With Early Preclinical Alzheimer's Disease and Intermediate Amyloid [Available from: https://ClinicalTrials.gov/show/NCT04468659.
13. Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. JAMA. 2015;313(19):1924-38.

14. Abi Nader C, Ayache N, Robert P, Lorenzi M, Alzheimer's Disease Neuroimaging I. Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data. Neuroimage. 2020;205:116266.
15. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 2013;12(4):357-67.
16. Mattsson N, Palmqvist S, Stomrud E, Vogel J, Hansson O. Staging beta-Amyloid Pathology With Amyloid Positron Emission Tomography. JAMA Neurol. 2019.
17. Palmqvist S, Scholl M, Strandberg O, Mattsson N, Stomrud E, Zetterberg H, et al. Earliest accumulation of beta-amyloid occurs within the default-mode network and concurrently affects brain connectivity. Nat Commun. 2017;8(1):1214.
18. Grothe MJ, Barthel H, Sepulcre J, Dyrba M, Sabri O, Teipel SJ, et al. In vivo staging of regional amyloid deposition. Neurology. 2017;89(20):2031-8.
19. Fantoni E, Collij L, Alves IL, Buckley C, Farrar G. The spatial-temporal ordering of amyloid pathology and opportunities for PET imaging. Journal of Nuclear Medicine. 2020;61(2):166-71.
20. Thal DR, Beach TG, Zanette M, Heurling K, Chakrabarty A, Ismail A, et al. [(18)F] flutemetamol amyloid positron emission tomography in preclinical and symptomatic Alzheimer's disease: specific detection of advanced phases of amyloid-beta pathology. Alzheimers Dement. 2015;11(8):975-85.
21. Bischof GN, Jacobs HIL. Subthreshold amyloid and its biological and clinical meaning: Long way ahead. Neurology. 2019;93(2):72-9.
22. Farrell ME, Chen X, Rundle MM, Chan MY, Wig GS, Park DC. Regional amyloid accumulation and cognitive decline in initially amyloid-negative adults. Neurology. 2018;91(19):e1809-e21.
23. Insel PS, Mormino EC, Aisen PS, Thompson WK, Donohue MC. Neuroanatomical spread of amyloid β and tau in Alzheimer's disease: Implications for primary prevention. Brain Communications. 2020.
24. Whittington A, Gunn RN, Alzheimer's Disease Neuroimaging I. Amyloid Load: A More Sensitive Biomarker for Amyloid Imaging. J Nucl Med. 2019;60(4):536-40.
25. Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, et al. Partial volume correction in quantitative amyloid imaging. Neuroimage. 2015;107:55-64.
26. Su Y, Flores S, Hornbeck RC, Speidel B, Vlassenko AG, Gordon BA, et al. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. Neuroimage Clin. 2018;19:406-16.
27. Chen K, Roontiva A, Thiyyagura P, Lee W, Liu X, Ayutyanont N, et al. Improved power for characterizing longitudinal amyloid-beta PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. J Nucl Med. 2015;56(4):560-6.

28. Chiao P, Bedell BJ, Avants B, Zijdenbos AP, Grand'Maison M, O'Neill P, et al. Impact of Reference and Target Region Selection on Amyloid PET SUV Ratios in the Phase 1b PRIME Study of Aducanumab. J Nucl Med. 2019;60(1):100-6.
29. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. J Nucl Med. 2017;58(7):1019-24.
30. Berckel B, Ossenkoppele R, Tolboom N, Yaqub M, Foster-Dingley J, Windhorst A, et al. Longitudinal amyloid imaging using [11C] PiB: methodological considerations. J Nucl Med. 2013.
31. Heeman F, Yaqub M, Alves IL, et al. Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies: J Cereb Blood Flow Metab. 2021;41:579-589.
32. Heeman F, Yaqub M, Lopes Alves I, Heurling K, Berkhof J, Gispert JD, et al. Optimized dual-time-window protocols for quantitative [(18)F]flutemetamol and [(18)F]florbetaben PET studies. EJNMMI Res. 2019;9(1):32.
33. Ottoy J, Verhaeghe J, Niemantsverdriet E, De Roeck E, Wyffels L, Ceyssens S, et al. (18) F-FDG PET, the early phases and the delivery rate of (18)F-AV45 PET as proxies of cerebral blood flow in Alzheimer's disease: Validation against (15)O-H2O PET. Alzheimers Dement. 2019;15(9):1172-82.
34. Bilgel M, Beason-Held L, An Y, Zhou Y, Wong DF, Resnick SM. Longitudinal evaluation of surrogates of regional cerebral blood flow computed from dynamic amyloid PET imaging. J Cereb Blood Flow Metab. 2020;40(2):288-97.
35. LaMontagne PJ, Benzinger TLS, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. medRxiv. 2019:2019.12.13.19014902.
36. Tolboom N, Yaqub M, Boellaard R, Luurtsema G, Windhorst AD, Scheltens P, et al. Test-retest variability of quantitative [11C]PiB studies in Alzheimer's disease. Eur J Nucl Med Mol Imaging. 2009;36(10):1629-38.
37. Su Y, D'Angelo GM, Vlassenko AG, Zhou G, Snyder AZ, Marcus DS, et al. Quantitative analysis of PiB-PET with FreeSurfer ROIs. PLoS One. 2013;8(11):e73377.
38. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. Neuroimage. 2009;46(1):154-9.
39. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968-80.
40. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. J Cereb Blood Flow Metab. 1996;16(5):834-40.

41. Svarer C, Madsen K, Hasselbalch SG, Pinborg LH, Haugbøl S, Frøkjær VG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. NeuroImage. 2005;24(4):969-79.
42. Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp. 2003;19(4):224-47.
43. Boellaard R, Yaqub M, Lubberink M, Lammertsma A. PPET: A software tool for kinetic and parametric analyses of dynamic PET studies. NeuroImage. 2006;31.
44. Collij LE, Heeman F, Salvado G, Ingala S, Altomare D, Wilde Ad, et al. Multi-tracer model for staging cortical amyloid deposition using PET imaging. Neurology. 2020.
45. Amadoru S, Dore V, McLean CA, Hinton F, Shepherd CE, Halliday GM, et al. Comparison of amyloid PET measured in Centiloid units with neuropathological findings in Alzheimer's disease. Alzheimers Res Ther. 2020;12(1):22.
46. Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement. 2012;8(1):1-13.
47. Timmers T, Ossenkoppele R, Visser D, Tuncel H, Wolters EE, Verfaillie SC, et al. Test-retest repeatability of [(18)F]Flortaucipir PET in Alzheimer's disease and cognitively normal individuals. J Cereb Blood Flow Metab. 2019:271678X19879226.
48. Collij LE, Konijnenberg E, Reimand J, Kate MT, Braber AD, Alves IL, et al. Assessing Amyloid Pathology in Cognitively Normal Subjects Using (18)F-Flutemetamol PET: Comparing Visual Reads and Quantitative Methods. J Nucl Med. 2019;60(4):541-7.
49. Guo T, Dukart J, Brendel M, Rominger A, Grimmer T, Yakushev I, et al. Rate of beta-amyloid accumulation varies with baseline amyloid burden: Implications for anti-amyloid drug trials. Alzheimers Dement. 2018;14(11):1387-96.
50. Joie R, Ayakta N, Seeley WW, Borys E, Boxer AL, DeCarli C, et al. Multisite study of the relationships between antemortem [(11)C]PiB-PET Centiloid values and postmortem measures of Alzheimer's disease neuropathology. Alzheimers Dement. 2018.
51. Battle M, Buckley C, Smith A, Farrar G, Thal D, Molineuvo JL, et al. Comparison of Centiloid Scaling Values with Visual Read Assessment in a Pathology Verified Autopsy Cohort. 2019.
52. Bernard Hanseeuw VM, Laurence Dricot, Lisa Quenon, Jiri Cerman, Christopher Buckley, Gill Farrar, Adrian Ivanoiu, Renaud Lhommel. P61: Defining a Centiloid scale threshold predicting long-term progression to dementia in patients attending the memory clinic: An F18-Flutemetamol amyloid-PET study. Human Amyloid Imaging Conference. 2020:211-2.
53. Leuzy A, Chiotis K, Hasselbalch SG, Rinne JO, de Mendonca A, Otto M, et al. Pittsburgh compound B imaging and cerebrospinal fluid amyloid-beta in a multicentre European memory clinic study. Brain. 2016;139(Pt 9):2540-53.

54. Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med. 2014;6(228):228fs13.
55. Lim YY, Mormino EC, Alzheimer's Disease Neuroimaging I. APOE genotype and early beta-amyloid accumulation in older adults without dementia. Neurology. 2017;89(10):1028-34.
56. Mishra S, Blazey TM, Holtzman DM, Cruchaga C, Su Y, Morris JC, et al. Longitudinal brain imaging in preclinical Alzheimer disease: impact of APOE epsilon4 genotype. Brain. 2018;141(6):1828-39.
57. Lopresti BJ, Campbell EM, Yu Z, Anderson SJ, Cohen AD, Minhas DS, et al. Influence of apolipoprotein-E genotype on brain amyloid load and longitudinal trajectories. Neurobiol Aging. 2020;94:111-20.
58. Ba M, Kong M, Li X, Ng KP, Rosa-Neto P, Gauthier S. Is ApoE varepsilon 4 a good biomarker for amyloid pathology in late onset Alzheimer's disease? Transl Neurodegener. 2016;5:20.
59. Mintun MA, Lo AC, Duggan Evans C, Wessels AM, Ardayfio PA, Andersen SW, et al. Donanemab in Early Alzheimer's Disease. N Engl J Med. 2021.
60. Moussa-Pacha NM, Abdin SM, Omar HA, Alniss H, Al-Tel TH. BACE1 inhibitors: Current status and future directions in treating Alzheimer's disease. Med Res Rev. 2020;40(1):339-84.
61. Affleck AJ, Sachdev PS, Stevens J, Halliday GM. Antihypertensive medications ameliorate Alzheimer's disease pathology by slowing its propagation. Alzheimers Dement (N Y). 2020;6(1):e12060.
62. Martorell AJ, Paulson AL, Suk HJ, Abdurrob F, Drummond GT, Guan W, et al. Multi-sensory Gamma Stimulation Ameliorates Alzheimer's-Associated Pathology and Improves Cognition. Cell. 2019;177(2):256-71 e22.
63. Rosenberg A, Mangialasche F, Ngandu T, Solomon A, Kivipelto M. Multidomain Interventions to Prevent Cognitive Impairment, Alzheimer's Disease, and Dementia: From FINGER to World-Wide FINGERS. J Prev Alzheimers Dis. 2020;7(1):29-36.
64. McNamee RL, Yee SH, Price JC, Klunk WE, Rosario B, Weissfeld L, et al. Consideration of optimal time window for Pittsburgh compound B PET summed uptake measurements. J Nucl Med. 2009;50(3):348-55.
65. Yaqub M, Tolboom N, Boellaard R, van Berckel BN, van Tilburg EW, Luurtsema G, et al. Simplified parametric methods for [11C]PiB studies. Neuroimage. 2008;42(1):76-86.
66. Veronese M, Bodini B, Garcia-Lorenzo D, Battaglini M, Bongarzone S, Comtat C, et al. Quantification of [(11)C]PiB PET for imaging myelin in the human brain: a test-retest reproducibility study in high-resolution research tomography. J Cereb Blood Flow Metab. 2015;35(11):1771-82.
67. Lopes Alves I, Collij LE, Altomare D, Frisoni GB, Saint-Aubert L, Payoux P, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. Alzheimers Dement. 2020;16(5):750-8.

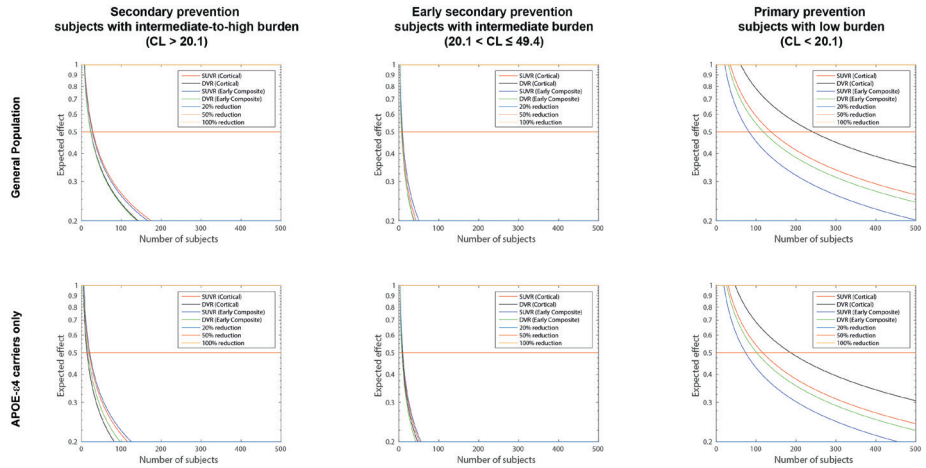
Supplementary methods

Using the LME estimates for annualized accumulation rates and respective standard deviations, the *sampsizepwr* function in Matlab (1-β = 80% power and a two-tailed t-test type-I error of α=0.05) was used to determine sample sizes required to detect a range of differences in accumulation rates (0-100% reduction) for a fixed power (80%). The overall trial design assumes participants undergo a PET scan at baseline and another at the completion of the trial. These were computed separately for SUVR and DVR, using the Cortical Composite and the Early Composite, both across the whole population and restricted to *APOE*-ε4 carriers only.

Supplementary results

As per main manuscript, the sample sizes (per arm) were computed for a primary prevention (CL ≤ 20.1), and two secondary prevention trials, one in subjects with intermediate amyloid burden (20.1 < CL ≤ 49.4), and the other in subjects with intermediate-to-high amyloid burden (CL > 20.1).

The overall effects described in the main manuscript are also observed here, where SUVR requires smaller sample sizes for primary prevention, while DVR provides better statistical power for secondary prevention designs. By estimating sample sizes as a function of expected effect, it is also noticeable that these differences are less relevant at larger effects, where sample sizes are similarly small for both metrics (Supplementary Figure 1).



Supplementary Figure 1. Relationship between sample size (per arm) and expected reductions in accumulation rates Proportional to baseline rates, i.e. y-axis values correspond to proportional reduction, where 0.2 is 20% reduction, 0.5 is 50%, and 1.0 is a complete halt in accumulation.

11.

Summary & General discussion

Summary

The overall aim of this thesis was to evaluate to what extent approaches for quantification of Aβ pathology can be simplified, within the context of different applications.

The main findings were that, in case of [^{18}F]flutemetamol, [^{18}F]florbetaben and [^{11}C]PiB, the semi-quantitative standardized uptake value ratio (SUVR) showed a bias that was proportional to underlying Aβ pathology compared with both the quantitative non-displaceable binding potential (BP_{ND}) and the distribution volume ratio (DVR). This suggests that for highly accurate quantification of Aβ pathology, such as in longitudinal studies, dynamic or validated dual-time window scanning protocols may be required. More specifically, the present work showed that dual-time window protocols can substitute dynamic protocols with only limited loss of accuracy, while increasing patient comfort and throughput. Clearly, highly accurate and robust quantification can only be achieved if these scanning protocols are used in combination with a validated pharmacokinetic model and reference tissue, if applicable. Furthermore, it was shown that bias in SUVR relative to BP_{ND} or DVR was dependent on the specific characteristics of the study population, and appeared to be smallest in participants with very low levels of Aβ pathology. Moreover, considering that Aβ plaque accumulation starts focally, research aiming to measure changes in Aβ pathology over time or relate pathology to cognitive performance in such a cohort, should focus on regional instead of global Aβ pathology measures. This strategy can increase power for detecting small changes in amyloid accumulation and may aid detection of distinct risk profiles in terms of cognitive decline. Finally, it should be noted that dynamic and dual-time window scanning protocols have the advantage of providing a measure of relative tracer delivery (R_1), which could serve as an additional biomarker for measuring disease progression.

Chapter 2 provides an overview of quantification strategies implemented in the European multi-centre project "Amyloid imaging to prevent Alzheimer's disease (AMYPAD)", which consists of the Diagnostic and Patient Management Study (DPMS) and the Prognostic and Natural History Study (PNHS). Static scanning protocols are implemented in the DPMS, as it focusses on the use of amyloid PET scans in clinical routine, where presence or absence of Aβ plaques is assessed visually. In the PNHS, more fine-grained estimates of Aβ pathology are required to track and predict disease progression, which is the reason for implementing dual-time window scanning protocols. Considering that two different tracers, [^{18}F]flutemetamol and [^{18}F]florbetaben, are used

within the project, results need to be standardised before pooling. To this end, the Centiloid method is used, which standardises amyloid PET measurements between tracers by converting SUVR measurement to a Centiloid value on a scale from 0-100, where the extremes correspond to young controls and typical Alzheimer's disease (AD) dementia patients, respectively. Furthermore, given that both studies include several centres, acquisition protocols were harmonised beforehand to allow for pooling scans. However, as the PNHS also includes historical PET data from other cohorts, harmonisation efforts for these scans were limited to the use of post-processing tools to maximize quantitative comparability.

Quantification of the specific amyloid tracers used within AMYPAD requires dynamic scanning protocols of nearly 2 hours duration, hence, their implementation would result in low patient comfort and scanning efficiency. Dynamic data acquisitions using a dual-time window acquisition protocol would improve patient comfort and scanner throughput. Based on previously acquired clinical data, a dual-time window protocol, with an optimal balance between quantitative accuracy and scanning time, was identified for both tracers by means of a simulation study in **Chapter 3.** To this end, quantitative accuracy of dual-time window data with breaks of various durations was compared with accuracy of complete dynamic data. As expected, for both tracers, longer breaks were associated with increasing bias in the quantitative parameter of interest, BP_{ND}. More specifically, a break of more than 60 min resulted in a high number of outliers and significant bias in BP_{ND} for both tracers. Consequently, the dual-time window scanning protocol consisting of a first scan from 0-30 min post injection (p.i.) and a second one from 90-110 min p.i. was selected as the best trade-off between quantitative accuracy and scanning time. An additional advantage of this protocol is that the length of the break allows for interleaved scanning (i.e. in the break of the first participant, the first scan of the next participant can be performed).

Having defined the optimal scanning protocol, the next step was to determine the optimal model for quantification of dual-time window PET data for both tracers. Taking into account that Aβ pathology might not follow anatomical boundaries and more detailed spatial information can be extracted using voxel-wise approaches, **Chapter 4** focussed on determining the optimal parametric method. To this end, 46 scans from either [^{18}F]flutemetamol or [^{18}F]florbetaben were used. Model performance was evaluated with respect to three different aspects: 1) visual assessment of parametric images 2) ability to differentiate between Aβ-positive and Aβ-negative scans, and 3) quantitative

accuracy and precision with respect to a regional reference method, in this case reference Logan (RLogan). Across these aspects, multilinear reference tissue model (MRTM0) performed best, closely followed by SRTM2 for both tracers. In the end, SRTM2 was chosen as the preferred method for parametric imaging, because it has the advantage of also providing a measure of relative tracer delivery (R_1), which could also be a useful biomarker for measuring disease progression in AD. Recent interest in this parameter (R_1) led to further investigation of its methodological characteristics in **Chapter 5**, where test-retest data were used to determine variability of [^{11}C]PiB R_1. As expected, global R_1 was significantly lower in AD dementia patients compared with cognitively unimpaired (CU) participants. Test-retest variability was low across all regions (1.5-5.8%), with a slightly higher variability for smaller regions. Furthermore, a negligible mean difference was observed between test and retest R_1. These findings are encouraging and, together with existing literature, suggests that [^{11}C]PiB R_1 can be used reliably in both cross-sectional and longitudinal studies.

Having determined the optimal dual-time window scan protocol together with accompanying quantification methodology, the next section of this thesis focused on a better understanding of the effects of confounding factors on different (semi-) quantitative approaches. More specifically, in **Chapter 6** the impact of CBF changes and implementation errors on quantitative and semi-quantitative outcome measures for both [^{18}F]flutemetamol and [^{18}F]florbetaben were evaluated. To this end, time-activity curves (TACs) were simulated using the reversible two-tissue compartment model with additional blood volume fraction parameter (2T4k_V_b) and CBF changes were simulated by varying tracer delivery (K_1) from +25 to -25% in a global cortical and/or cerebellar reference region. SUVR was calculated across a range of uptake times and quantitative estimates of Aβ burden were derived using the simplified reference tissue model (SRTM) and RLogan. Across methods, the extent of sensitivity to CBF generally increased with higher levels of Aβ burden, and RLogan was least affected by changes in CBF. Finally, effects of expedited or delayed uptake times on quantification were small, although in general, SUVR's sensitivity to CBF changes was smaller for later uptake times. Taken together, these findings demonstrate that from these three methods, RLogan would be the preferred method for regional quantification of Aβ pathology for both tracers. Building onto this simulation work, **Chapter 7** investigated whether SUVR's bias relative to DVR can be explained by factors such as relative CBF (as measured by relative tracer delivery, R_1) and underlying Aβ burden. To this end, 121 scans from the PNHS were used, focusing on early amyloid burden and containing both [^{18}F]flutemetamol and [^{18}F]florbetaben

scans. For both tracers, good correlations were observed between SUVR and DVR, although bias between the two metrics was strongly related to underlying Aβ burden, with increasing bias for higher levels of Aβ pathology. In contrast to findings in AD dementia patients from a previous [^{11}C]PiB study, bias in SUVR could not be explained by R_1 effects. This may be explained by the fact that the vast majority of participants were cognitively unimpaired and changes in CBF are more pronounced in clinical AD populations. Next, **Chapter 8** focussed on one specific violation of the reference tissue approach, namely the choice of a reference tissue with different characteristics than the target tissue. In recent years, several studies have proposed alternative reference tissues to the cerebellum grey matter, mostly for improving effect sizes when measuring Aβ pathology longitudinally. Test-retest (N=13) and longitudinal (N=30) datasets were used to make a comprehensive evaluation of their performance. Four reference regions, whole cerebellum, white matter brainstem/pons, whole brainstem and eroded subcortical white matter, were used to derive quantitative and semi-quantitative outcome measures for [^{11}C]PiB, and these results were compared with the gold standard, i.e. plasma-input based quantification using cerebellum grey matter. Across various aspects, whole and grey matter cerebellum reference regions provided best results compared with the gold standard, and they were therefore considered optimal for quantification of Aβ pathology using [^{11}C]PiB PET.

Following assessment of the impact of confounding factors on (semi-)quantitative measures of Aβ pathology, the last section of this thesis focussed on advantages of applying regional and quantitative approaches in a clinical or clinical trial setting. Standard methods for determining amyloid abnormality are dichotomous assessments based on a visual read or global SUVR, which carries the danger that the earliest signs of Aβ pathology may be missed. Recent work has demonstrated the clinical value of identifying multiple stages in terms of Aβ pathology, but it is unclear whether these models can be generalised across tracers. **Chapter 9** proposes a multi-tracer model for staging cortical regional Aβ pathology using data from CU participants. This model assumes that regions most frequently abnormal in a CU population reflect early events in the pathologic process, which was tested on 4783 scans acquired using four different amyloid tracers. The data-driven model detected abnormal Aβ plaque levels most frequently in cingulate, followed by orbitofrontal, precuneal and insular cortices and subsequently in associative, temporal and occipital regions. The model showed high applicability across amyloid tracers and cohorts with participants across the AD spectrum. Analyses using DVR and partial volume corrected SUVR data resulted in

comparable models, demonstrating that the effect of confounders and spill-in effects from white matter on the proposed ordering was limited. Furthermore, the staging model could be used for detecting pre-global Aβ pathology and identify distinct risk profiles in terms of cognitive decline. This suggests that moving away from dichotomised measures of Aβ pathology by evaluating regional Aβ burden has the potential to improve participant selection for clinical trials and predict cognitive decline in a clinical setting. Finally, **Chapter 10** evaluated strategies for improving statistical power in AD prevention trials using [^{11}C]PiB PET for detecting subtle-to-moderate biomarker changes in three different trial scenarios. These included a primary prevention trail scenario, which aims to prevent AD, as well two secondary prevention trials, which aim to reduce the impact of the AD, following its onset. In this work, tabulated PET data of 237 participants from the Open Access Series of Imaging Studies (OASIS-3) dataset were included together with a test-retest dataset from 11 subjects. Results showed that sample sizes in secondary AD prevention trials can be reduced by acquisition of dynamic (as opposed to static) scans and restricting inclusion criteria to *APOE*-ε4 carriers only or inclusion of subjects with intermediate Aβ burden. With respect to primary AD prevention trials, use of a targeted early composite region for Aβ quantification proved to be a viable strategy for reducing sample sizes as well as restricting inclusion to *APOE*-ε4 carriers.

General discussion

The perfect trade-off?

Over the past decades, there have been ongoing efforts to simplify full tracer kinetic modelling procedures, which rely on arterial blood sampling and dynamic PET acquisitions. Around the turn of the century, multiple reference tissue models have been introduced to circumvent the burdensome procedure of blood sampling, and these methods are now routinely used for research purposes (*1–6*). More recently, it has been investigated whether long dynamic acquisition could be made more patient friendly, while retaining high quantitative accuracy, by introducing two shorter acquisitions (*7–10*). For multiple tracers, it has now been demonstrated that dual-time window protocols can be accurate and more patient friendly alternatives for full dynamic scanning protocols (*7,11*). In particular, in this thesis (Chapter 3), optimal dual-time window protocols for [^{18}F]flutemetamol and [^{18}F]florbetaben were identified and validated using simulations that were based on clinical data. In addition, it was demonstrated that this protocol offers the additional benefit of enabling interleaved scanning, which can

improve efficiency in terms of tracer batch and scanner usage and thereby reduce costs. However, there are also some practical issues that need to be taken into account. First, two scans are needed for attenuation correction purposes, which, in the case of a PET/CT scanner, adds to the total amount of radiation exposure of the patient. Furthermore, dual-time window scans require more demanding processing steps compared with full dynamic acquisitions, as early and late parts of the scan need to be coregistered, and interpolation of missing data points of the reference tissue curve is required to allow for pharmacokinetic modelling. Considering the advantages of dual-time window scanning protocols, it is not surprising that these protocols have become more popular for research purposes (*8–10*). Nevertheless, because of the higher demand in terms (pre-)processing steps, this type of acquisition protocol has primarily been used at experienced PET centres. Although knowledge of processing and pharmacokinetic modelling will remain essential for defining optimal model settings and configurations, it would be valuable to the PET community at large to have a straightforward, single tool that incorporates both (pre-)processing and quantification steps necessary for analyses of these PET data, thereby facilitating analysis of large datasets and improving consistency amongst researchers (*12*). Currently, efforts are undertaken within the Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) consortium to create such a tool.

Effects of underlying Aβ pathology on the relationship between SUVR and DVR

Despite the recently growing number of scans acquired using a dual-time window protocol, in most clinical and research settings a static protocol remains the standard for measuring Aβ pathology, mainly because of its technical and logistical simplicity. However, the outcome measure from a static scan (standardized uptake value ratio, SUVR) is known to be biased compared with the distribution volume ratio (DVR) or non-displaceable binding potential (BP_{ND}), derived from a dynamic scan (*13*). In particular, it has been shown that the bias between these metrics may change as a function of underlying pathology (*14,15*). To date, some studies have compared simplified and fully quantitative metrics at a diagnostic group level. However, as disease stage and underlying Aβ pathology are not linearly related, a clear understanding of how bias changes as function of Aβ pathology is still missing (*16–18*). In this thesis, this relationship was evaluated for [^{18}F]flutemetamol, [^{18}F]florbetaben (Chapter 4 and 7) and [^{11}C]PiB (Chapter 10) and, in line with theory, it was demonstrated that for each of these tracers, SUVR overestimated DVR and that this overestimation was directly associated with the underlying Aβ pathology (*14*). More specifically, overestimation was lowest in participants with very low Aβ pathology and highest in participants with extensive Aβ

pathology. While this could suggest that routine usage of static scans in participants in early Alzheimer's disease (AD) stages would be acceptable, some examples might help illustrate the challenges with such an approach.

The shifted focus of clinical trials to preclinical (i.e. before manifestation of clinical symptoms) stages of AD, implies more participants with low Aβ burden will be included in those trials (*19*). For trial inclusion, presence of Aβ plaques usually is based on a visual read of a static scan. However, it has been demonstrated that, especially in individuals with low Aβ burden, a clinical visual read of an SUVR image will result in a higher number of false-positives compared with a visual read of a BP_{ND} image, thus possibly resulting in inappropriate trial inclusion (*15*). Although including semi-quantitative measures (i.e. SUVR) in addition to the visual assessments of the SUVR images reduces such false-positive classifications, these can be eliminated by using a BP_{ND} images (*15*). Another example is related to the proportional overestimation of DVR by SUVR, resulting in artificially inflated SUVR-based accumulation rates in longitudinal studies (i.e. studies with multiple measurements over a period of time). As demonstrated in a primary prevention trial scenario in this thesis, inflated accumulation rates facilitated detection of treatment effects when using SUVR compared with DVR in participants with low Aβ burden (Chapter 10). Consequently, a treatment might be considered more effective than it actually is. Finally, the difference in bias between regions with low and extensive Aβ pathology can have implications for between group comparisons, given that the measured difference will be inflated when using SUVR, which may be particularly problematic if small differences are expected (*14*). Taken together, these examples illustrate that in particular for applications where measuring small changes or small group differences is of interest, DVR may be required. On the other hand, for applications where large changes or large group differences are expected and the exact magnitude of the difference is trivial, SUVR may be sufficient.

Impact of CBF on bias in SUVR

In addition to the effect of Aβ pathology on the relationship between SUVR and DVR, it is important to also consider the impact of CBF on this relationship, given that estimation of SUVR does not take into account (and correct for) other physiological changes, such as reductions in CBF, which are characteristic for AD dementia (*20*,*21*). Such CBF reductions may prolong the time needed for the tracer to reach equilibrium conditions, which will affect SUVR to a certain extent, depending on the tracer's kinetics (*22*). The impact of CBF reductions on SUVR was illustrated by a longitudinal [^{11}C]PiB

study, which demonstrated that in a group of AD dementia patients, CBF reductions (as measured by R_1) were most likely the cause of concomitant reductions in SUVR, while no reductions in DVR were observed (*13*). However, this effect was not observed in mild cognitive impaired (MCI) patients, despite the fact that several studies reported CBF reductions already in MCI stages (*23–25*). Furthermore, Sojkova and colleagues reported differences between regional changes in CBF over time (i.e. longitudinal), as measured with [^{15}O]H_2O PET, in high and low Aβ burden groups, in a cohort of mostly cognitively unimpaired (CU) participants (*26*). The possible interaction between CBF and Aβ pathology and its impact on quantification, resulted in investigating the impact of CBF changes on SUVR and DVR for different levels of Aβ pathology by means of a simulation study (Chapter 6). This work demonstrated that reference Logan (RLogan)-derived DVR was least sensitive to changes in CBF (compared with simplified reference tissue model (SRTM)-derived DVR and SUVR) and this sensitivity generally increased with increasing levels of Aβ pathology. This means that in regions with low Aβ pathology, time to reach equilibrium is relatively independent of CBF. On the other hand, in regions with extensive Aβ pathology more time is required to reach equilibrium, which can be expedited or delayed by CBF in or decreases.

A follow-up study using clinical data investigated whether (bias in) SUVR (relative to DVR) could be explained by CBF, as measured by R_1, in a cohort of mostly CU individuals. Although negative associations were observed between (bias in) SUVR and CBF for both [^{18}F]flutemetamol and [^{18}F]florbetaben, these results were not significant (Chapter 7). A plausible explanation for this finding is that changes in CBF were likely too small in this study population (*27*). In any case, these results indicate that CBF is likely not to be a major source of bias for SUVR in mainly CU individuals. Some methodological considerations should also be taken into account, given that the relative tracer delivery parameter R_1 is only a proxy of CBF. First, given that R_1 is a relative measure, it cannot be used to measure global changes (i.e. in target as well as reference tissues) in CBF. However, in the simulation work (Chapter 6) it was demonstrated that the impact of global changes in CBF on SUVR was comparable to that of regional changes in terms of magnitude. Second, as opposed to the gold standard [^{15}O]H_2O PET CBF, R_1 is also affected by the extraction fraction of the tracer, which suggests it may be a less sensitive metric for detecting subtle changes in CBF, as was demonstrated recently (*28*). Therefore, for future studies, it would be valuable to measure both CBF and SUVR in a longitudinal setting and to investigate the impact of CBF as measured with [^{15}O]H_2O PET on bias in SUVR, to further elucidate the interaction between Aβ pathology and

CBF. Finally, when interpreting these findings, it is important to note that the studies discussed above were observational studies, which means that changes in CBF were related only to aging or disease related processes (*29*). In interventional designs (i.e. clinical trials), the drugs being tested may also affect CBF and usually these effects are poorly understood prior to the start of the study. Importantly, if large CBF changes are present, these may seriously compromise the trials' results, at least when using SUVR as outcome measure. Therefore, it should be considered to use dynamic acquisitions to avoid this issue.

Sources of bias for simplified approaches

In addition to physiological effects of Aβ pathology and CBF, there are other factors that can compromise the use of simplified approaches for quantification. As discussed above, one of the first simplifications to the procedure of full tracer kinetic modelling was the introduction of reference tissue models to circumvent the need for arterial blood sampling by providing an alternative input function. This alternative input function, the tracers' activity over time in a reference tissue, is derived from a region devoid of specific binding sites, but with similar tissue characteristics as the target tissue, which should be unaffected by the disease under investigation. In the case of imaging Aβ pathology in AD, cerebellar grey matter (CBGM) fulfils these criteria in nearly all patients, and has therefore commonly been used for this purpose (*30*,*31*). In recent years, alternative reference regions have been proposed, as CBGM may become implicated in very late stages or familial forms of AD dementia (*32*–*34*), or aiming to provide higher effect sizes for measuring Aβ accumulation. However, many of these proposed alternative reference regions consists of or include white matter, thus violating the assumption of similar tissue characteristics between target and reference tissues. Despite this assumption violation, several studies have used white matter reference regions and reported higher effect sizes for detecting longitudinal changes in amyloid burden (*35*–*37*). As can be observed in Chapter 8, these findings could be related to the lower test-retest variability (i.e. higher precision) for some regions consisting (partially) of white matter tissue, as has also been suggested previously (*38*). More specifically, in Chapter 8 it was demonstrated that the whole cerebellum appeared to be the region of choice in terms of stability, while CBGM provided more comparable results with the gold standard (i.e. DVR, calculated indirectly by dividing the target tissue volume of distribution (V_T) by the V_T of the CBGM reference tissue), in the case of [^{11}C]PiB. Thus, one's choice of reference region may depend on which aspect is important from a research question perspective. However, the higher effect sizes obtained with reference regions consisting mainly of white matter

tissue should be interpreted with caution, as age related white matter increases have been reported and disease related changes are still poorly understood (*38*). In addition, knowledge of non-specific white matter binding of amyloid tracers is still limited (*39*,*40*).

Another aspect that should be considered when using reference tissue models is related to the specific set of assumptions of each of them. For example for the SRTM, a simulation study characterized the bias that can be introduced by violating these assumptions (*41*). For [^{18}F]flutemetamol and [^{18}F]florbetaben used within AMYPAD, it has been shown that both target and reference region kinetics are better described by a two-tissue compartment model (*42*,*43*), thus violating SRTMs' assumption of single tissue kinetics in both regions (*2*). In line with these findings, Chapter 3 demonstrated a small bias in non-linear SRTM-derived DVR when fitting 2T4k_V_b simulated [^{18}F]flutemetamol or [^{18}F]florbetaben time-activity curves (TACs) and Chapter 6 reported poorly determined DVR and k_2 values due to poor fits of 2T4k_V_b simulated TACs with CBF changes, which were both likely a result of the assumption violation. These findings illustrate the importance of using a pharmacokinetic model that is valid for describing the kinetics of the data, in order to obtain reliable quantitative results. It should be noted that differences in outcome measures between pharmacokinetic models can also be more subtle as reported for model comparisons in Chapter 4. Here, reference tissue models were compared based on the following aspects: 1) visual assessment of parametric images 2) ability to differentiate between Aβ-positive and Aβ-negative scans, and 3) quantitative accuracy and precision with respect to a regional reference tissue method, in this case RLogan. Across aspects, the multilinear reference tissue model (MRTM0) performed best, closely followed by SRTM2 for both tracers. In the end, SRTM2 was chosen as the preferred method for parametric imaging, because it has the advantage of also providing a measure of relative tracer delivery. The optimal performance of SRTM2 compared with the other reference tissue models, likely reflects the need for fewer fitting parameters and the more robust estimate of k_2', which both reduce SRTM2's sensitivity to noise. Furthermore, adequate performance of both receptor parametric mapping (RPM) and SRTM2 in Chapter 4, in view of the poor performance of SRTM on simulated data, is likely due to the effect of basis functions, which limit the number of possible model solutions and thereby prevent fit-parameters from becoming undetermined. For some applications, such as when determining whether there is a difference between diagnostic groups, these small between-model differences in performance may be negligible, while for other applications, such as when tracking yearly disease progression, these differences

could be highly relevant. Thus, it may depend on the research question, which model will be preferred.

Last, it is important to note that in clinical and research practise, logistical errors, i.e. deviations in terms of delayed or expedited static scan start times, occur regularly. Previous work demonstrated that these deviations impact accuracy of SUVR measurement (*13*), and Chapter 6 showed that these deviations also affect SUVR's sensitivity to CBF changes. More specifically, the difference in SUVR's sensitivity to CBF changes was max. 5% across uptake times, suggesting that for small deviations from the recommended uptake window (max. 10 min), the effect on SUVR would be negligible.

Application of amyloid-PET – moving away from binary classifications

In addition to the technical aspects concerning measurement of Aβ pathology using PET, this thesis also focussed on optimizing the use of amyloid PET scans for clinical and research purposes. Traditionally, AD research has focussed mainly on late diseases stages, characterised by cognitive impairment. In recent years, however, the focus has shifted towards earlier disease stages, characterized by abnormal biomarker levels (*44–47*). A simple binary classification of the PET scan has been routinely used for both clinical and research purposes, but this approach neglects two major advantage of amyloid PET, namely its topographical and quantitative information. In particular, it has been demonstrated that already in CU participants, Aβ pathology starts focally and can range from very little to full blown pathology (*27,48*). Thus, especially in these individuals, the use of binary classification tools for assessing Aβ pathology may miss the earliest signs of pathology (*49*). Recent studies have shown that using regional approaches, instead of binary ones, may provide insight in both the spatial distribution of Aβ pathology as well as its extent, as illustrated by a study that reported high correlations between the number of visual Aβ-positive regions and the extent of Aβ pathology as measured in Centiloid (CL) units (*50*). Moreover, moving away from a global, binary classification also has clinical relevance, as shown by the potential value of regional Aβ pathology and its accumulation rates for predicting cognitive decline (*51,52*). In addition, it has been reported that utilizing the quantitative aspect of amyloid PET (i.e. evaluating the extent of Aβ pathology) can predict progression to AD dementia in non-demented memory clinic patients based on pre-defined CL cut-off values (*53*). This predictive information may be valuable for decision making in clinical routine to determine who could benefit from anti-amyloid treatment, such as with the recently approved aducanumab (Aduhelm) (*54*). Work in this thesis extended these findings by demonstrating that

stages of regional Aβ pathology in globally Aβ-positive individuals were related to distinct risk profiles of cognitive decline (Chapter 9). In addition, these stages showed relationships both with established AD biomarkers such as A*POE*-ε4 carriership, cerebral spinal fluid (CSF) Aβ42, CSF tau and with clinical diagnosis. Furthermore, in agreement with other studies, the staging model, constructed with data from four amyloid PET tracers, indicated that cingulate, orbitofrontal and precuneal regions were most frequently abnormal in terms of amyloid PET uptake, in CU individuals (*55–57*). This information may be particularly relevant to studies aiming to detect Aβ plaques or Aβ accumulation using PET in individuals with low overall Aβ pathology (*55*). In fact, this is illustrated in Chapter 10, where using early regions for detecting changes in Aβ accumulation rates in individuals with low Aβ burden (CL ≤ 20.1) reduced the required sample size with 40-50%. This finding is likely because accumulation rates were diluted (i.e. weakened) when using a large, global region, as a result of the inclusion of both affected and non-affected regions. Together, the aforementioned findings indicate that going beyond binary classifications of amyloid PET scans may harbour relevant information for both research as well as clinical practice. Specifically, with clinical AD trials moving into prevention strategies were low levels of Aβ pathology are expected, it has been suggested that it might become essential to evaluate Aβ pathology before global positivity is reached for inclusion purposes, warranting regional or quantitative approaches (*58*). In addition, these trials might need to take into account both the extent of pathology and its accumulation rates for determining the effectiveness of an intervention (*58*). Nevertheless, a better understanding of the prognostic value of this early Aβ pathology can serve to further guide clinical trial design, which is what the AMYPAD prognostic and natural history study (PNHS) aims to contribute to (*58*). In addition, this prognostic information will be useful to determine which patients could benefit from treatment in the clinical setting.

Future perspectives

Multi-centre and multi-tracer studies

The work presented in this thesis was part of the European, multicentre AMYPAD project, which meant that PET data from many different centres were included. This brought analysis challenges due to e.g. between-centre differences in scanner, data dimensions, formats, and naming conventions. This also meant that existing software tools used for the analysis of dual-time window scans first had to be validated for data

from each individual centre. Such extensively validated tools are not only an asset to AMYPAD itself, but also for future projects. Another issue regularly encountered in multi-centre studies, is that of multiple tracers and cross-tracer comparability of quantitative measures. Recently, the Centiloid method has been introduced to allow for cross-tracer comparability of SUVR, by converting amyloid SUVR measurements to a Centiloid value between 0-100, with the two extremes corresponding to Aβ pathology levels in young controls and typical AD dementia patients, respectively (*59*) (Chapter 2). To date, such an approach is not yet available for quantitatively comparing DVR values from different amyloid tracers. However, the growing number of dynamic scans collected with various amyloid tracers emphasizes the need for a standardized measure to allow for pooling these data for meta-analyses. Therefore, within AMYPAD, efforts for creating a standardized DVR measure are currently ongoing. Until such a standardized measure has been developed, possibilities for comparing DVR across tracers will be limited to binary comparisons (i.e. normal versus abnormal) using tracer specific thresholds. As demonstrated in Chapter 9, such thresholds could be calculated in a data-driven manner or, alternatively, be defined against other established AD biomarkers (*60*). In addition to cross-tracer comparability of DVR, there is also the question of cross-tracer comparability of R_1. In particular considering its recent increase in popularity as a proxy for relative cerebral blood flow and metabolism, it is important to understand whether findings can be extrapolated across tracers (*28*,*61*–*63*). This question was outside the scope of the current thesis and remains to be addressed in future studies.

What's next for quantitative amyloid imaging?

Despite the critical note on the more burdensome acquisition and analysis procedures of dual-time window protocols, use of those protocols has increased during the last decade and could be an interesting option for other tracers in the future. In addition to providing an unbiased and specific estimate of Aβ pathology, another advantage of dual-time window scans is that a proxy of relative CBF (e.g. R_1) can be calculated from the early frames. This is particularly relevant for AD research, considering that reductions in CBF are characteristic for AD dementia and have also been reported in MCI stages (*21*,*25*). Moreover, promising findings in terms of moderate-to-high cross-sectional correlations between [^{11}C]PiB-derived R_1 and [^{15}O]H_2O-derived CBF have been reported (*64*). Considering that R_1 has only gained attention recently, studies investigating its longitudinal properties are still scarce (*28*). In this thesis, Chapter 5 contributed to understanding its longitudinal value, by investigating the intrinsic variability of R_1 in order to determine what magnitude of change in R_1 signifies an actual change.

Furthermore, a better characterisation of how (change in) this metric relates to other AD related biomarkers and cognitive decline is warranted to fully understand its potential for clinical and research purposes (*63*). Nonetheless, results from Chapter 7 suggested that in the earliest stages of AD, changes in (relative) CBF are still limited. Therefore, in these stages, it may be more relevant/informative to focus on early biomarkers such as Aβ pathology (*27*).

In terms of the application of amyloid PET, it has been demonstrated that in particular in individuals in early AD phases, extracting regional and quantitative as opposed to dichotomised measures of Aβ pathology can provide valuable information regarding its spatial distribution and the extent of pathology. These findings may guide studies aiming to detect accumulation in individuals with low Aβ burden and potentially have predictive clinical value (*51*,*52*). Such predictive information will be particularly relevant for selecting participants for both primary and secondary prevention trials and to determine who will be eligible for receiving anti-amyloid treatment in a diagnostic setting. In addition, using quantitative estimates of Aβ pathology may also be important to evaluate the effectiveness of interventions with only small effect sizes. Given that the shifted research-focus to early AD phases has only taken place recently, there are still numerous questions related to the prognostic value of regional and quantitative Aβ pathology that need to be addressed, in which the AMYPAD PNHS may play an important role. In particular, AMYPAD's purpose to provide an open access dataset at the end of the study duration will facilitate exploration of these questions by researchers around the world.

Conclusion

This thesis evaluated to what extent approaches for quantification of Aβ pathology can be simplified within the context of different applications. It has been demonstrated that simplifications may come at a price and defining the right balance between accuracy and simplicity can be a delicate issue, as many different factors need to be considered. Nonetheless, one should always start with the clinical or research question that needs to be addressed and then consider, after fully characterising simplifications by comparing them with fully quantitative studies, what simplifications are deemed appropriate.

References

1. Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. *J Cereb Blood Flow Metab*. 1996;16:834-840.
2. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. *Neuroimage*. 1996;4:153-158.
3. Ichise M, Liow J-S, Lu J-Q, et al. Linearized Reference Tissue Parametric Imaging Methods: Application to [11 C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human Brain. *Journal of Cerebral Blood Flow & Metabolism*. 2003;23:1096-1112.
4. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric Imaging of Ligand-Receptor Binding in PET Using a Simplified Reference Region Model. *NeuroImage*. 1997;6:279-287.
5. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. *J Cereb Blood Flow Metab*. 2002;22:1440-1452.
6. Cunningham VJ, Hume SP, Price GR, Ahier RG, Cremer JE, Jones AKP. Compartmental Analysis of Diprenorphine Binding to Opiate Receptors in the Rat in vivo and its Comparison with Equilibrium Data in vitro. *Journal of Cerebral Blood Flow & Metabolism*. 1991;11:1-9.
7. Bullich S, Barthel H, Koglin N, et al. Validation of Noninvasive Tracer Kinetic Analysis of 18F-Florbetaben PET Using a Dual-Time-Window Acquisition Protocol. *J Nucl Med*. 2018;59:1104-1110.
8. Cecchin D, Barthel H, Poggiali D, et al. A new integrated dual time-point amyloid PET/MRI data analysis method. *European Journal of Nuclear Medicine and Molecular Imaging*. 2017;44:2060-2072.
9. Florek L, Tiepolt S, Schroeter ML, et al. Dual Time-Point [18F]Florbetaben PET Delivers Dual Biomarker Information in Mild Cognitive Impairment and Alzheimer's Disease. *J Alzheimers Dis*. October 2018.
10. Son SH, Kang K, Ko P-W, et al. Early-Phase 18F-Florbetaben PET as an Alternative Modality for 18F-FDG PET. *Clinical Nuclear Medicine*. 2020;45:e8.
11. Tuncel H, Visser D, Yaqub M, et al. Effect of Shortening the Scan Duration on Quantitative Accuracy of [18F]Flortaucipir Studies. *Mol Imaging Biol*. January 2021.
12. Veronese M, Rizzo G, Belzunce M, et al. Reproducibility of findings in modern PET neuroimaging: insight from the NRM2018 grand challenge. *J Cereb Blood Flow Metab*. May 2021:271678X211015101.
13. Berckel van BNM, Ossenkoppele R, Tolboom N, et al. Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. *J Nucl Med*. 2013;54:1570-1576.

14. Carson RE, Channing MA, Blasberg RG, et al. Comparison of Bolus and Infusion Methods for Receptor Quantitation: Application to [18 F]Cyclofoxy and Positron Emission Tomography. *Journal of Cerebral Blood Flow & Metabolism*. 1993;13:24-42.
15. Collij LE, Konijnenberg E, Reimand J, et al. Assessing Amyloid Pathology in Cognitively Normal Subjects Using 18F-Flutemetamol PET: Comparing Visual Reads and Quantitative Methods. *J Nucl Med*. 2019;60:541-547.
16. Golla SS, Verfaillie SC, Boellaard R, et al. Quantification of [18F]florbetapir: A test–retest tracer kinetic modelling study. *J Cereb Blood Flow Metab*. June 2018:0271678X18783628.
17. Heurling K, Buckley C, Van Laere K, Vandenberghe R, Lubberink M. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18)F]flutemetamol. *Neuroimage*. 2015;121:184-192.
18. Becker GA, Ichise M, Barthel H, et al. PET Quantification of 18F-Florbetaben Binding to -Amyloid Deposits in Human Brains. *Journal of Nuclear Medicine*. 2013;54:723-731.
19. Eisai Inc. AHEAD 3-45 Study: A Placebo-Controlled, Double-Blind, Parallel-Treatment Arm, 216 Week Study to Evaluate Efficacy and Safety of Treatment With BAN2401 in Subjects With Preclinical Alzheimer's Disease and Elevated Amyloid (A45 Trial) and in Subjects With Early Preclinical Alzheimer's Disease and Intermediate Amyloid (A3 Trial). clinicaltrials.gov; 2021.
20. Binnewijzend MAA, Benedictus MR, Kuijer JPA, et al. Cerebral perfusion in the predementia stages of Alzheimer's disease. *Eur Radiol*. 2016;26:506-514.
21. Postiglione A, Lassen NA, Holman BL. Cerebral blood flow in patients with dementia of Alzheimer's type. *Aging Clin Exp Res*. 1993;5:19-26.
22. Lammertsma AA. Forward to the Past: The Case for Quantitative PET Imaging. *J Nucl Med*. 2017;58:1019-1024.
23. Johnson NA, Jahng GH, Weiner MW, et al. Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: Initial experience. *RADIOLOGY*. 2005;234:851-859.
24. Collij LE, Heeman F, Kuijer JPA, et al. Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer Disease. *Radiology*. 2016;281:865-875.
25. Binnewijzend MAA, Kuijer JPA, Benedictus MR, et al. Cerebral blood flow measured with 3D pseudocontinuous arterial spin-labeling MR imaging in Alzheimer disease and mild cognitive impairment: a marker for disease severity. *Radiology*. 2013;267:221-230.
26. Sojkova J, Beason-Held L, Zhou Y, et al. Longitudinal Cerebral Blood Flow and Amyloid Deposition: An Emerging Pattern? *Journal of Nuclear Medicine*. 2008;49:1465-1471.
27. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol*. 2010;9:119-128.

28. Bilgel M, Beason-Held L, An Y, Zhou Y, Wong DF, Resnick SM. Longitudinal evaluation of surrogates of regional cerebral blood flow computed from dynamic amyloid PET imaging. *J Cereb Blood Flow Metab*. 2020;40:288-297.
29. Frackowiak RS. Measurement and imaging of cerebral function in ageing and dementia. *Prog Brain Res*. 1986;70:69-85.
30. Catafau AM, Bullich S, Seibyl JP, et al. Cerebellar Amyloid-β Plaques: How Frequent Are They, and Do They Influence 18F-Florbetaben SUV Ratios? *Journal of Nuclear Medicine*. 2016;57:1740-1745.
31. Klunk WE, Wang Y, Huang G, et al. The binding of 2-(4'-methylaminophenyl)benzothiazole to postmortem brain homogenates is dominated by the amyloid component. *J Neurosci*. 2003;23:2086-2092.
32. Thal DR, Rüb U, Orantes M, Braak H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology*. 2002;58:1791-1800.
33. Wegiel J, Wisniewski HM, Dziewiatkowski J, et al. Cerebellar atrophy in Alzheimer's disease—clinicopathological correlations. *Brain Research*. 1999;818:41-50.
34. Lomoio S, López-González I, Aso E, et al. Cerebellar amyloid-β plaques: disturbed cortical circuitry in AβPP/PS1 transgenic mice as a model of familial Alzheimer's disease. *J Alzheimers Dis*. 2012;31:285-300.
35. Schwarz CG, Senjem ML, Gunter JL, et al. Optimizing PiB-PET SUVR Change-Over-Time Measurement by a large-scale analysis of Longitudinal Reliability, Plausibility, Separability, and Correlation with MMSE. *Neuroimage*. 2017;144:113-127.
36. Fleisher AS, Joshi AD, Sundell KL, et al. Use of white matter reference regions for detection of change in florbetapir positron emission tomography from completed phase 3 solanezumab trials. *Alzheimers Dement*. 2017;13:1117-1124.
37. Su Y, Blazey TM, Owen CJ, et al. Quantitative Amyloid Imaging in Autosomal Dominant Alzheimer's Disease: Results from the DIAN Study Group. Herholz K, ed. *PLoS ONE*. 2016;11:e0152082.
38. Lowe VJ, Lundt ES, Senjem ML, et al. White Matter Reference Region in PET Studies of 11C-Pittsburgh Compound B Uptake: Effects of Age and Amyloid-β Deposition. *J Nucl Med*. 2018;59:1583-1589.
39. Sebai S, Baciu M, Ces O, et al. To lipophilicity and beyond—towards a deeper understanding of radioligand non-specific binding. *NeuroImage*. 2006;31:T56.
40. Fodero-Tavoletti MT, Rowe CC, McLean CA, et al. Characterization of PiB Binding to White Matter in Alzheimer Disease and Other Dementias. *Journal of Nuclear Medicine*. 2009;50:198-204.
41. Salinas CA, Searle GE, Gunn RN. The simplified reference tissue model: model assumption violations and their impact on binding potential. *J Cereb Blood Flow Metab*. 2015;35:304-311.

42. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 Study of the Pittsburgh Compound B Derivative 18F-Flutemetamol in Healthy Volunteers and Patients with Probable Alzheimer Disease. *Journal of Nuclear Medicine*. 2009;50:1251-1259.
43. Becker GA, Ichise M, Barthel H, et al. PET quantification of 18F-florbetaben binding to β-amyloid deposits in human brains. *J Nucl Med*. 2013;54:723-731.
44. Ingala S, Boer CD, Masselink LA, et al. Application of the ATN classification scheme in a population without dementia: Findings from the EPAD cohort. *Alzheimer's & Dementia*. n/a.
45. Sperling RA, Donohue MC, Raman R, et al. Association of Factors With Elevated Amyloid Burden in Clinically Normal Older Individuals. *JAMA Neurol*. 2020;77:735-745.
46. van der Kall LM, Truong T, Burnham SC, et al. Association of β-Amyloid Level, Clinical Progression, and Longitudinal Cognitive Change in Normal Older Individuals. *Neurology*. 2021;96:e662-e670.
47. Sakr FA, Grothe MJ, Cavedo E, et al. Applicability of in vivo staging of regional amyloid burden in a cognitively normal cohort with subjective memory complaints: the INSIGHT-preAD study. *Alzheimer's Research & Therapy*. 2019;11:15.
48. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. *Neurology*. 2017;89:2031-2038.
49. Fantoni E, Collij L, Lopes Alves I, Buckley C, Farrar G, AMYPAD consortium. The Spatial-Temporal Ordering of Amyloid Pathology and Opportunities for PET Imaging. *J Nucl Med*. 2020;61:166-171.
50. Collij LE, Salvadó G, Shekari M, et al. Visual assessment of [18F]flutemetamol PET images can detect early amyloid pathology and grade its extent. *Eur J Nucl Med Mol Imaging*. February 2021.
51. Farrell ME, Chen X, Rundle MM, Chan MY, Wig GS, Park DC. Regional amyloid accumulation and cognitive decline in initially amyloid-negative adults. *Neurology*. 2018;91:e1809-e1821.
52. Hanseeuw BJ, Betensky RA, Mormino EC, et al. PET staging of amyloidosis using striatum. *Alzheimers Dement*. 2018;14:1281-1292.
53. Hanseeuw BJ, Malotaux V, Dricot L, et al. Defining a Centiloid scale threshold predicting long-term progression to dementia in patients attending the memory clinic: an [18F] flutemetamol amyloid PET study. *Eur J Nucl Med Mol Imaging*. 2021;48:302-310.
54. Research C for DE and. FDA's Decision to Approve New Treatment for Alzheimer's Disease. *FDA*. June 2021.
55. Insel PS, Mormino EC, Aisen PS, Thompson WK, Donohue MC. Neuroanatomical spread of amyloid β and tau in Alzheimer's disease: implications for primary prevention. *Brain Commun*. 2020;2:fcaa007.

56. Palmqvist S, Schöll M, Strandberg O, et al. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat Commun*. 2017;8:1214.
57. Mattsson N, Palmqvist S, Stomrud E, Vogel J, Hansson O. Staging β-Amyloid Pathology With Amyloid Positron Emission Tomography. *JAMA Neurol*. 2019;76.
58. Bischof GN, Jacobs HIL. Subthreshold amyloid and its biological and clinical meaning: Long way ahead. *Neurology*. 2019;93:72-79.
59. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia*. 2015;11:1-15.e4.
60. Salvadó G, Molinuevo JL, Brugulat-Serrat A, et al. Centiloid cut-off values for optimal agreement between PET and CSF core AD biomarkers. *Alz Res Therapy*. 2019;11:27.
61. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. 18F-FDG PET, the early phases and the delivery rate of 18F-AV45 PET as proxies of cerebral blood flow in Alzheimer's disease: Validation against 15O-H2O PET. *Alzheimer's & Dementia*. 2019;15:1172-1182.
62. Peretti DE, Vállez García D, Reesink FE, et al. Relative cerebral flow from dynamic PiB scans as an alternative for FDG scans in Alzheimer's disease PET studies. *PLoS One*. 2019;14:e0211000.
63. Visser D, Wolters EE, Verfaillie SCJ, et al. Tau pathology and relative cerebral blood flow are independently associated with cognition in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. May 2020.
64. Chen YJ, Rosario BL, Mowrey W, et al. Relative 11C-PiB Delivery as a Proxy of Relative CBF: Quantitative Evaluation Using Single-Session 15O-Water and 11C-PiB PET. *J Nucl Med*. 2015;56:1199-1205.

Addendum

Nederlandse Samenvatting

Ziekte van Alzheimer

Als gevolg van de toenemende vergrijzing van de bevolking, neemt ook de incidentie van dementie toe. Alleen al in Nederland lijden er meer dan 280.000 mensen aan dementie en de meest voorkomende oorzaak hiervan (±70% van de gevallen) is de ziekte van Alzheimer. Deze ziekte wordt gekenmerkt door progressieve cognitieve klachten, zoals geheugenklachten en problemen bij het uitvoeren van executieve functies. Onderzoek heeft uitgewezen dat de eerste pathologische veranderingen in de hersenen al twee decennia voor de klinische presentatie van de ziekte kunnen plaatsvinden. De meest kenmerkende pathologische veranderingen zijn de opeenhoping van de eiwitfragmenten amyloïde en tau, hetgeen uiteindelijke atrofie, ofwel hersenkrimp, tot gevolg heeft. De voornaamste hypothese binnen het wetenschappelijk onderzoek naar de ziekte van Alzheimer, is dat de opeenhoping van amyloïde (ook wel amyloïde "plaques" genoemd) de eerste pathologische veranderingen in de hersenen is. Zodoende is aanwezigheid van deze amyloïde plaques in de hersenen een zeer geschikte biomarker voor de ziekte van Alzheimer.

Amyloïde PET-beeldvorming

Sinds het begin van deze eeuw is het mogelijk om het amyloïde eiwit in beeld te brengen en de hoeveelheid ervan te kwantificeren met behulp van positron emissie tomografie (PET). PET-beeldvorming maakt gebruik van minuscule hoeveelheden radioactief-gelabelde stoffen (tracers), die een interactie aangaan met (in dit geval) het amyloïde eiwit in het brein. Bij het verval van het radioactieve deel van de tracer komt gammastraling vrij die wordt gedetecteerd door de PET-detectoren. Op basis van deze informatie kan er een 3D beeld van de verdeling van het eiwit in de hersenen worden gereconstrueerd. Ook kunnen meerdere metingen over de tijd gedaan worden om zodoende een 4D beeld te reconstrueren dat het volledige proces van opname, verspreiding en aanhechting van de tracer vastlegt. In dit proefschrift is onderzoek gedaan met vier verschillende tracers voor visualisatie en kwantificatie van het amyloïde eiwit, namelijk: [^{11}C]PiB, [^{18}F]flutemetamol, [^{18}F]florbetaben en [^{18}F]florbetapir.

Momenteel wordt PET-beeldvorming in de klinische praktijk gebruikt om visueel te beoordelen of amyloïde plaques aanwezig zijn bij patiënten waarbij een vermoedden bestaat op de ziekte van Alzheimer. Voor het stellen van deze diagnose is de aanwezigheid van deze plaques een vereiste, alleen zijn de plaques op zich niet voldoende. Amyloïde

plaques komen namelijk ook bij asymptomatische, cognitief gezonde ouderen voor, waarvan nog onbekend is of ze ooit de ziekte van Alzheimer zullen ontwikkelen. Zodoende is het belangrijk om meer inzicht te krijgen in de rol die de opeenhoping van het amyloïde eiwit speelt bij het ontstaan van deze ziekte. Derhalve is het van groot belang om het eiwit in de hersenen over verloop van tijd volgen, liefst beginnend in de preklinische fase van de ziekte. Met behulp van deze informatie kan er hopelijk beter voorspeld worden wie een verhoogd risico heeft op het ontwikkelen van de ziekte van Alzheimer. Vervolgens kunnen deze risico indicatoren gebruikt worden om te bepalen wie er in aanmerking zouden moeten komen voor bestaande medicijnen of wie er deel zouden moeten nemen aan studies naar de werking van nieuwe medicijnen.

Kwantificatie van het amyloïde eiwit met PET

Voor nauwkeurige bepalingen van het amyloïde eiwit is kwantificatie een vereiste. Hiervoor is een 4D beeld, ofwel een dynamische scan nodig. Een dynamische scan bestaat uit een serie in de tijd opeenvolgende 3D PET-beelden. In tegenstelling tot statische scans waarbij alleen een semi-kwantitatieve maat (de standardized uptake value ratio, SUVR) berekend kan worden, kunnen dynamische scans veranderingen in andere fysiologische processen, zoals perfusie, meenemen in het verkrijgen van een specifieke maat van de hoeveelheid amyloïde in het brein, de non-displaceable binding potential (BP_{ND}). Voor deze berekening zijn niet alleen de PET-beelden nodig, maar ook een geschikt wiskundig (farmacokinetisch) model en arteriële monsters om de hoeveelheid radioactiviteit in het bloedplasma vast te stellen. Aangezien het hier om complexe, invasieve (arteriële monsters) en tijdrovende procedures gaat, is er de afgelopen jaren veel aandacht besteed aan het vereenvoudigen van de benodigde procedures. Zowel het inkorten van de scan acquisitie tijd (deze kan oplopen tot bijna 2 uur voor een dynamische amyloïde PET-scan), alsmede het vermijden van arteriële monsters zijn twee van de meest onderzochte vereenvoudigingen. Binnen de nucleaire geneeskunde wordt veel gebruik gemaakt van deze vereenvoudigde methodes. Echter, voor implementatie van deze methodes, is het belangrijk dat ze gevalideerd worden ten opzichte van de gouden standaardprocedure (d.w.z. een dynamische scan met arteriële monsters). Aangezien de kinetiek van elke tracer anders is, moet deze validatie voor iedere nieuwe tracer herhaald worden. Desalniettemin, ook na validatie van vereenvoudigde methodes ten opzichte van de gouden standaardprocedure, zullen er situaties zijn waarin deze vereenvoudigingen niet toereikend zijn.

Het Europese AMYPAD-consortium

Binnen het Europese Amyloid imaging to prevent Alzheimer's disease (AMYPAD) consortium worden verschillende aspecten van de diagnostische en prognostische waarde van amyloïde PET-scans onderzocht binnen twee sub-studies, de diagnostic and patient management study (DPMS) en de prognostic and natural history study (PNHS). Het werk in dit proefschrift is onderdeel van het AMYPAD-project en valt onder de prognostische sub-studie.

Het voornaamste doel van dit proefschrift was het vaststellen van de mate waarin kwantificatie van amyloïde plaques vereenvoudigd kan worden met het oog op verschillende toepassingen. De hoofdstukken zijn onderverdeeld in drie thema's: kwantificatie van amyloïde PET-scans, het effect van diverse factoren op deze kwantificatie en tot slot de impact van kwantificatie op zowel klinische praktijk als medicijnstudies.

Kwantificatie van amyloïde PET-scans

In **hoofdstuk 2** wordt een overzicht gegeven van de verschillende procedures die zijn gebruikt voor kwantificatie van amyloïde PET-scans binnen het Europese, multicenter AMYPAD-project. Binnen de diagnostische studie werden er statische scans gemaakt om de studie zo goed mogelijk te laten aansluiten bij het gebruik van PET in de kliniek. In deze studie is de primaire uitkomstmaat een dichotome, visuele beoordeling van de scan. In de prognostische studie werd een deel van de deelnemers longitudinaal gevolgd om het verloop van de ziekte en de stapeling van amyloïde plaques te monitoren. Zodoende werd er binnen deze studie veelal dynamisch gescand aan de hand van een dual-time window protocol (d.w.z. een protocol bestaande uit twee kortere scans) met als uitkomst maat de hoeveelheid specifiek gebonden tracer aan amyloïde plaques, BP_{ND}. Aangezien er binnen AMYPAD twee amyloïde PET-tracers zijn gebruikt, [^{18}F]flutemetamol en [^{18}F]florbetaben, werd de gestandaardiseerde Centiloid schaal als uitkomstmaat gebruikt om resultaten van beide tracers samen te kunnen voegen. Daarnaast wordt er in dit hoofdstuk aandacht besteed aan harmonisatie van het PET-acquisitie protocol tussen centra, alsmede aan harmonisatie van bestaande scans van andere cohorten die achteraf werden opgenomen in het AMYPAD-project. Gezien het lange dynamische acquisitie protocol (110 minuten) voor zowel [^{18}F]flutemetamol als [^{18}F]florbetaben, met als gevolg een hoge patiënt belasting en een lage efficiëntie m.b.t. tracer en scanner gebruik, werd er in **hoofdstuk 3** gekeken of dit acquisitie protocol ingekort kon worden. Op basis van eerdere, veelbelovende resultaten richtte de in dit hoofdstuk beschreven studie zich op het valideren van het

optimale dual-time window protocol, waarbij de acquisitie tijd werd verkort door het inlassen van een pauze tussen het eerste en tweede deel van de scan. Door middel van een simulatiestudie op basis van klinische data werd het effect van een pauze van variërende lengte op de kwantitatieve uitkomstmaat voor specifieke binding aan amyloïde (BP_{ND}, berekend met het simplified reference tissue model, SRTM) onderzocht. Dit onderzoek wees uit dat een pauze van 60 minuten optimaal was met betrekking tot inkorten van de scan enerzijds en nauwkeurigheid van de meting anderzijds. Dit resulteerde in een dual-time window protocol bestaande uit een eerste scan van 0-30 min, 60 min pauze en een tweede scan van 90-110 min na injectie. Daarbij heeft dit protocol als voordeel dat de eerste scan van de tweede patiënt tijdens de pauze van de eerste patiënt gemaakt kan worden. Nadat het optimale scan protocol gedefinieerd was, was het belangrijk het optimale model voor kwantificatie van deze scans te definiëren. Aangezien er meer gedetailleerde spatiële informatie uit PET-scans verworven kan worden door gebruik te maken van parametrische ofwel voxel-gebaseerde modellen (een voxel is een 3D pixel), was het doel van **hoofdstuk 4** het optimale parametrische model voor kwantificatie van dual-time window [^{18}F]flutemetamol en [^{18}F]florbetaben scans te definiëren door diverse parametrische methodes te vergelijken met een bestaande regionale referentie methode. Zesenveertig scans van de AMYPAD-PNHS werden geïncludeerd en de prestatie van zeven parametrische modellen werd beoordeeld aan de hand van drie aspecten, 1) visuele beoordeling van het parametrische beeld, 2) het vermogen om onderscheid te maken tussen positieve en negatieve amyloïde scans en 3) nauwkeurigheid en precisie ten opzichte van een regionale referentie methode (reference Logan, RLogan). Het multilinear reference tissue model (MRTM0) liet de beste prestaties zien over de verschillende aspecten, direct gevolgd door SRTM2. SRTM2 werd daarop geselecteerd als meest geschikte model voor de kwantificatie van deze scans, omdat m.b.v. SRTM2, naast BP_{ND}, ook een maat van relatieve cerebrale perfusie (relative tracer delivery, R_1) verkregen kan worden. Deze maat, R_1, kan namelijk als additionele biomarker voor Alzheimer-gerelateerde ziekteprogressie gebruikt worden. Positieve eerdere bevindingen m.b.t. het gebruik van R_1 als maat voor relatieve cerebrale perfusie heeft geleid tot een toename in het gebruik hiervan, ook in longitudinale studies. Om vast te stellen of een verandering in R_1 een daadwerkelijke verandering is of slechts variabiliteit van de meting, werd in **hoofdstuk 5** de variabiliteit van deze maat voor [^{11}C]PiB onderzocht op basis van een bestaande test-hertest dataset. Zoals verwacht was R_1 lager in patiënten met Alzheimer's dementie dan in gezonde deelnemers. Daarnaast was er heel weinig variatie in de R_1 meting, n.l. 1,7% voor een globale, corticale hersenregio en 1,5-5,8%

voor kleinere hersenregio's. Kortom, deze bevindingen ondersteunen het gebruik van R_1 in zowel cross-sectioneel als longitudinaal onderzoek.

Het effect van diverse factoren op kwantificatie

Nadat het optimale acquisitie protocol en het bijbehorende parametrische model voor kwantificatie waren gedefinieerd, was de volgende stap het in kaart brengen van de invloed van diverse mogelijk verstorende factoren (confounding factors) op de gouden standaard en de daarvan afgeleide vereenvoudigde maten.

Een veel gerapporteerde fysiologische verandering in de klinische fase van de ziekte van Alzheimer is een afname in de cerebrale perfusie (doorbloeding). Daarnaast hebben enkele studies ook relatieve toenames gerapporteerd in verschillende delen van de cortex ten opzichte van het cerebellum. Zodoende werd in **hoofdstuk 6** het effect van veranderingen in de cerebrale perfusie (zowel toe- als afname) op kwantificatie van BP_{ND} onderzocht aan de hand van een simulatiestudie. Drie verschillende soorten veranderingen in cerebrale perfusie werden gesimuleerd: 1) veranderingen in het gehele brein, d.w.z. in zowel een globale corticale hersenregio als het cerebellum (de kleine hersenen), 2) veranderingen in de globale corticale hersenregio en 3) veranderingen in het cerebellum. Vervolgens werd de hoeveelheid amyloïde plaques op drie manieren gekwantificeerd, n.l. met twee farmacokinetische modellen (SRTM en RLogan) en met SUVR berekend over het late (statische) deel van de scan. Bij de laatste methode werd ook onderzocht of er een interactie was tussen het verschuiven van de starttijd van de statische scan en de sensitiviteit van SUVR voor veranderingen in cerebrale perfusie. De simulaties toonden aan dat, over het algemeen, kwantificatie meer werd beïnvloed door veranderingen in cerebrale perfusie als er meer amyloïde plaques in het brein aanwezig waren en dat RLogan hier het minst gevoelig voor was. Het verschuiven van de scan starttijd had relatief weinig effect, echter, over het algemeen was SUVR minder gevoelig voor veranderingen in cerebrale perfusie bij latere scan starttijden. Uit dit werk vloeide **hoofdstuk 7** voort, waarin [^{18}F]flutemetamol en [^{18}F]florbetaben scans van de PNHS werden geïncludeerd (N=121) en werd onderzocht of de afwijking in SUVR ten opzichte van BP_{ND}+1 (hier distribution volume ratio, DVR genoemd) kon worden verklaard door de mate van onderliggende amyloïde pathologie of door veranderingen in relatieve cerebrale perfusie (gemeten door R_1). Ondanks hoge correlaties tussen SUVR en DVR, was er een substantiële proportionele overschatting van DVR door SUVR. Deze bevinding is met name belangrijk voor longitudinale studies, omdat deze afwijking ervoor kan zorgen dat de gemeten toename in SUVR hoger is

dan de daadwerkelijke toename in amyloïde pathologie. De afwijking in SUVR t.o.v. DVR kon niet worden verklaard door (veranderingen in) de relatieve cerebrale perfusie (R_1), waarschijnlijk doordat het merendeel van de deelnemers cognitief gezond was en grote veranderingen in cerebrale perfusie met name gerapporteerd zijn in klinische populaties met Alzheimer's dementie. Het is aannemelijk dat dit ook de reden is waarom eerder onderzoek een longitudinale afname in zowel SUVR als R_1 rapporteerde voor een groep patiënten met Alzheimer's dementie, terwijl dit niet gevonden werd voor gezonde deelnemers of deelnemers met milde cognitieve achteruitgang. Tot slot lag de focus van **hoofdstuk 8** op het gebruik van een alternatieve input voor arteriële monsters, namelijk de activiteit over de tijd in een hersenregio waarin geen specifieke binding van de tracer wordt verwacht: een z.g. referentie hersenregio. Eén van de aannames waaraan geschikte referentie hersenregio's moeten voldoen, is dat deze bestaan uit hetzelfde type weefsel als de hersenregio die onderzocht wordt (ook wel de "target" hersenregio genoemd). Echter, aangezien dit niet altijd het geval is, werd in dit hoofdstuk onderzocht in hoeverre (semi-)kwantificatie wordt beïnvloed als referentie hersenregio's hier niet (geheel) aan voldoen. Op basis van een test-hertest (N=13) en longitudinale (N=30) [^{11}C]PiB dataset werd het gebruik van alternatieve referentie hersenregio's vergeleken met resultaten verkregen met de gouden standaard, plasma-input kwantificatie, met de grijze stof van het cerebellum als referentie hersenregio. Er werd gebruik gemaakt van deze gevalideerde referentie regio om vergelijkbare eenheden (d.w.z. DVR) als met de andere methodes te krijgen. Daarnaast werd gekeken naar de variabiliteit van de meting en naar het vermogen om onderscheid te maken tussen positieve en negatieve amyloïde scans. De onderzochte alternatieve referentie hersenregio's waren: het gehele cerebellum, de witte stof van de hersenstam/pons, de gehele hersenstam en een geërodeerde subcorticale witte stof regio (d.w.z. waarbij voxels aan de rand van deze regio, grenzend aan de grijze stof, zijn verwijderd). Voor alle vijf referentie hersenregio's werd amyloïde pathologie gekwantificeerd met RLogan, SRTM en SUVR en de prestaties van de verschillende regio's werden beoordeeld op basis van de vergelijking met de gouden standaard, de variabiliteit van de meting en het vermogen onderscheid te maken tussen positieve en negatieve amyloïde scans. Op basis van de onderzochte aspecten lieten de grijze stof van het cerebellum en het gehele cerebellum voor de verschillende methodes de beste resultaten zien. Zodoende werden deze regio's aangemerkt als optimale referentie hersenregio's voor kwantificatie van amyloïde pathologie met [^{11}C]PiB.

Impact van kwantificatie op de klinische praktijk en medicijnstudies

Het laatste deel van het proefschrift richtte zich op het gebruik van amyloïde PET-scans in de klinische setting. Hierin werden de voordelen van PET-kwantificatie onderzocht t.o.v. de standaard, dichotome visuele beoordeling, die routinematig in de klinische praktijk wordt gebruikt. Een nadeel van de visuele beoordeling is dat de vroegste pathologische veranderingen in de hersenen makkelijk worden gemist. Eerder onderzoek toonde aan dat stadiëring van amyloïde pathologie op basis van een model dat de spatiële-temporele volgorde van de betrokkenheid van de verschillende hersenregio's in kaart brengt klinische relevantie heeft. Deze studies lieten zien dat hogere stadia (d.w.z. betrokkenheid van meer hersenregio's) werden geassocieerd met een verhoogd risico op cognitieve achteruitgang. Echter, in deze studies werden verschillende tracers en populaties bestudeerd, hetgeen leidde tot verschillen in de volgorde van de betrokken hersengebieden. Zodoende werd in **hoofdstuk 9** een overkoepelend stadieëringsmodel ontwikkeld o.b.v. data van vier verschillende amyloïde PET-tracers. Het model dat in dit hoofdstuk werd ontwikkeld, bestond uit vijf stadia en was in staat om meer dan 99% van de 4783 scans waarop het werd toegepast te classificeren. In overeenstemming met bestaande literatuur, werden onder andere de cingulate en orbitofrontale cortices alsmede de precuneus bestempeld als hersenregio's met vroege amyloïde stapeling. Daarnaast werden er in deze studie associaties aangetoond met andere klinische variabelen zoals diagnose, cerebrospinale eiwit waardes en met het *APOE*-ε4 gen dat een verhoogd genetisch risico op het ontwikkelen van de ziekte van Alzheimer met zich meebrengt. Ten slotte liet dit hoofdstuk zien dat een stadieëringsmodel potentie heeft om amyloïde pathologie te detecteren voordat de scan op basis van een dichotome beoordeling als positief zou worden aangemerkt en dat het model verschillende risicoprofielen voor cognitieve achteruitgang kon identificeren. Deze bevindingen m.b.t. de voorspellende waarde van regionale stadiëringsmodellen suggereren dat ze mogelijk van belang kunnen zijn in het beter selecteren van proefpersonen voor medicijnstudies en in het voorspellen wie er in aanmerking zou moeten komen voor anti-amyloïde medicijnen, zoals het recent goedgekeurde medicijn aducanumab (Aduhelm). **Hoofdstuk 10** richtte zich specifiek op medicijnstudies en hierin werd onderzocht wat de toegevoegde waarde was van het gebruik van een regionale amyloïde maat (bestaande uit hersenregio's met vroege amyloïde stapeling) en dynamische PET-scans voor de detectie van subtiele veranderingen in amyloïde pathologie. In deze studie werd gebruik gemaakt van de OASIS [^{11}C]PiB dataset, bestaande uit 237 deelnemers. Drie hypothetische medicijnstudie scenario's werden onderzocht waarbij in ieder het doel was om de toename van amyloïde opstapeling met 20% te verminderen. De power-analyses lieten zien dat

het gebruik van dynamische PET-scans met de daarbij horende kwantitatieve maat (DVR) tot een substantiële vermindering van het aantal benodigde proefpersonen kon leiden in secundaire preventiestudies. Daarnaast kon inclusie van proefpersonen met intermediaire amyloïde belasting het aantal benodigde proefpersonen met een factor vier verkleinen. In primaire preventiestudies lag het aantal benodigde proefpersonen een stuk hoger, maar dit aantal kon met 40-50% omlaag worden gebracht door het kwantificeren van amyloïde pathologie in vroege hersengebieden of door alleen dragers van het *APOE*-ε4 gen in de studie te includeren.

Conclusie

In dit proefschrift werd onderzocht in welke mate procedures voor het kwantificeren van amyloïde plaques vereenvoudigd kunnen worden in het kader van verschillende toepassingen. Ook werd er aandacht besteed aan het effect van verschillende verstorende factoren op kwantificatie. Er werd aangetoond dat vereenvoudigde methodes de nauwkeurigheid van de bepaling van de hoeveelheid amyloïde plaques in het brein kunnen beïnvloeden. Zodoende is het cruciaal dat deze vereenvoudigde methodes eerst worden vergeleken met de gouden standaard, zodat in kaart kan worden gebracht wat het effect hiervan is op kwantificatie. Op basis van deze informatie en met inachtneming van de beoogde toepassing, kan worden besloten of en zo ja welke vereenvoudigingen gerechtvaardigd zijn.

List of abbreviations

Aβ	Amyloid-β
ABIDE	Alzheimer's Biomarkers in Daily practice study
AC	Attenuation correction
AIC	Akaike Information Criterion
AD	Alzheimer's disease
ADC	Alzheimer's Dementia Cohort
ADNI	Alzheimer's disease Neuroimaging Initiative
ALFA	Alzheimer's and Family cohort
AMYPAD	Amyloid Imaging to Prevent Alzheimer's disease
ANOVA	One-way analysis of variance
***APOE*-ε4**	Apolipoprotein-ε4
AUC	Area under the curve
BA	Bland-Altman
BL	Baseline
BP_{ND}	Non-displaceable binding potential
Bq	Becquerel
CBF	Cerebral blood flow
CDR	Clinical and Dementia Rating scale
C_F	Free concentration in tissue
CSF	Cerebrospinal fluid
CL	Centiloid
CI	Confidence interval
CN	Cognitively normal
C_{NS}	Non-specifically bound concentration in tissue
COV	Coefficient of Variation
C_P	Tracer radioactivity concentration in plasma
C_S	Specifically bound concentration in tissue
CT	Computed Tomography
CU	Cognitively unimpaired
C-11/^{11}C	Carbon -11
DK	Desikan-Kiliany atlas
DPMS	Diagnostic and Patient Management Study
DVR	Distribution Volume Ratio
EANM	European Association of Nuclear Medicine

EARL	EANM Research Ltd.
EMA	European Medicine Agency
EPAD LCS	European Prevention of Alzheimer's Dementia Longitudinal Cohort Study
FDA	Food and Drug Administration
FDG	Fluoro-2-deoxy-glucose
FOV	Field of view
FRTM	Full Reference Tissue Model
FU	Follow-up
F-18 /^{18}F	Fluorine-18
GAAIN	Global Alzheimer's Association Interactive Network
GCA	Global cortical average
GLM	Generalised linear models
GMCB	Grey matter cerebellum
HC	Healthy control
HR	Hazard ratio
IQR	Interquartile range
IDM	Iterative deconvolution method
keV	Kiloelectron volt
KM	Kaplan-Meier
K_1	Delivery rate constant
k_2 / k_3 / k_4	Rate constants
LB	Lower boundary
LEAP	Learning Embeddings for Atlas Propagation
LME	Linear Mixed Effect Models
LMM	Linear Mixed Model
LoA	Limits of Agreement
LOR-RAMLA	Standard line-of-response–based row-action maximum-likelihood algorithm
MAD	Median Absolute Deviation
MBq	Megabecquerel
MCI	Mild cognitive impairment
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
MRTM	Multilinear Reference Tissue Model
NEMA	National Electrical Manufacturers Association

NIfTI	Neuroimaging Informatics Technology Initiative
OASIS	Open Access Series of Imaging Studies
OFC	Orbitofrontal cortex
PCC	Posterior cingulate cortex
PET	Positron Emission Tomography
PI	Plasma input
p.i.	Post injection
PiB	Pittsburgh Compound B
PNHS	Prognostic and Natural History Study
PVC	Partial volume correction
PVE	Partial volume effect
QC	Quality control
RLogan	Reference Logan
ROC	Receiver Operating Characteristic
ROI	Region of interest
RPM	Receptor Parametric Mapping
RR	Reference Region
R_1	Relative tracer delivery rate
SD	Standard deviation
SE	Standard error
SRTM	Simplified Reference Tissue Model
SRTM2	Simplified Reference Tissue Model 2
SUV	Standardized Uptake Value
SUVR	Standardized Uptake Value Ratio
SCD	Subjective cognitive decline
TAC	Time-activity curve
TRT	Test-retest
2T4k_V_b	Two-tissue compartment model with four rate constants and additional blood volume fraction parameter
UB	Upper boundary
UCL	University College London
VOI	Volume of interest
VR	Visual read
V_{ND}	Non-displaceable volume of distribution
V_T	Volume of distribution
WBS	Whole brainstem

WCB	Whole cerebellum
WMBS	White matter brainstem
WMES	Eroded subcortical white matter

About the author

Fiona was born in Biddinghuizen, Flevoland on the 3rd of February 1992. She lived there until she was 18 and moved to Groningen to study Psychology. Besides her studies, she was very actively involved in many other aspects of student-life through participation in committees, boards and councils. During her studies, she realized that in order to unravel human behaviour, she needed a more comprehensive understanding of human biology and the brain. Combined with her drive to explore new places, she decided to move to Flagstaff, Arizona, USA to study abroad for one semester. There, she had the opportunity to attend a variety of courses in the field of biology, neurology and pharmacology to fulfil this need. Following her stay abroad, she attended bachelor and master courses in the field of Neuropsychology. However, her mind was set; she wanted to obtain a master's degree in Neurosciences at the VU University Amsterdam. During the master's degree, she became acquainted with neuroimaging techniques and completed her first internship at the department of Radiology and Nuclear Medicine of the VUmc in 2015. During a second year neuroimaging course, she was introduced to positron emission tomography (PET), which immediately sparked her interest and eventually resulted in her moving to London in 2016 for half a year, to study epilepsy patients using PET at UCLH. This internship convinced her that she wanted to learn more about PET, which eventually led to her starting a PhD position at the Amsterdam UMC, location VUmc with prof. dr. Adriaan Lammertsma.

During her PhD, she attended multiple international conferences, supervised student projects and published several scientific papers. In addition to her scientific work, she was involved in organising the ONWAR annual meeting for Neuroscience PhD students and co-organised the institute Quantivision conference *Machine Learning in Medical Imaging*. Following her 4-year PhD trajectory, she continued working on the AMYPAD project as postdoctoral researcher. The next step in her career will be a 2-year position as postdoctoral researcher at the University of Gothenburg in Sweden, starting shortly after her defence.

PORTFOLIO

Courses	Date	ECTS
Basiscursus klinisch onderzoekers (BROK)	2017	1.50
ONWAR Introductory course	2017	0.90
Working with radioactivity- official level 5B	2017	3.00
Neuropsychopharmacology	2017	1.40
MATLAB Fundamentals	2018	1.43
MATLAB Data Processing and Visualisation	2018	0.30
PET Pharmacokinetic course	2018	0.90
Writing a Data Management Plan	2018-2019	1.00
Scientific Integrity	2019	2.00
Functional Neuroanatomy	2019	1.40
Grant writing	2020	1.80
Workshops and National conferences		
Europond Disease Modelling	2017	0.25
ONWAR annual meeting: poster and blitz presentation	2017-2020	3.60
Amsterdam Neurosciences annual meeting: poster presentation	2018-2020	2.00
Brain PET Educational course	2019	0.30
International conferences		4.00
AAIC (London)	2017	
IMI 10th Anniversary Scientific Symposium (Brussels): oral presentation	2018	
Neuro Receptor Mapping (London): poster presentation	2018	
Human Amyloid Imaging (Miami): poster presentation	2019	
Brain & Brain PET (Yokohama): oral presentation	2019	
Human Amyloid Imaging (Miami): poster presentation	2019	
EANM: e-poster	2020	
Human Amyloid Imaging (Miami): poster presentation	2021	
AAIC (Denver, virtual): 2 poster presentations	2021	
Other academic activities		
ONWAR Career Event	2017	0.30
ONWAR Swammerdam Lectures	2017-2020	0.71
Active participation in research meetings and seminars	2017-2021	2.00
ONWAR annual meeting: organizing committee	2018-2019	5.00
iQ Conference Machine Learning in Medical Imaging: organizing committee	2019	
Teaching		2.00
Guest lecturer: Pharmacokinetics Master course	2019	
Mentoring 1 BSc and 1 MSc student	2019-2020	
Lecture: Functional Brain Imaging Master course	2019-2020	
Lecture: Medical Oncology PET course	2020	
Awards		
Early Career Investigator Travel Bursary Brain & Brain PET	2019	
Travel Fellowship AAIC	2021	

List of Publications

Collij, L. E., **Heeman, F.**, Kuijer, J. P. A., Ossenkoppele, R., Benedictus, M. R., Möller, C., Verfaillie, S. C. J., Sanz-Arigita, E. J., van Berckel, B. N. M., van der Flier, W. M., Scheltens, P., Barkhof, F. & Wink, A. M. (2016). Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer Disease. Radiology, 281(3), 865–875. https://doi.org/10.1148/radiol.2016152703

Heeman, F., Yaqub, M., Lopes Alves, I., Heurling, K., Berkhof, J., Gispert, J. D., Bullich, S., Foley, C. & Lammertsma, A. A. O*n behalf of the AMYPAD consortium.* (2019). Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. EJNMMI Research, 9(1), 32.

Heeman, F., Yaqub, M., Lopes Alves, I., Heurling, K., Bullich, S., Gispert, J. D., Boellaard, R. & Lammertsma, A. A. O*n behalf of the AMYPAD consortium.* (2021). Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies: Journal of Cerebral Blood Flow & Metabolism, 41(3), 579–589. https://doi.org/10.1177/0271678X20918029

Heeman, F.,* Collij, L. E.*, Salvadó, G., Ingala, S., Altomare, D., de Wilde, A., Konijnenberg, E., van Buchem, M., Yaqub, M., Markiewicz, P., Golla, S. S. V., Wottschel, V., Wink, A. M., Visser, P. J., Teunissen, C. E., Lammertsma, A. A., Scheltens, P., van der Flier, W. M., Boellaard, R., van Berckel, B.N.M., Molinuevo, J.L., Gisper, J.D., Schmidt., M.E., Barkhof, F. & Lopes Alves, I. *For the ALFA Study, For the Alzheimer's Disease Neuroimaging Initiative,* O*n behalf of the AMYPAD consortium* (2020). Multitracer model for staging cortical amyloid deposition using PET imaging. Neurology, 95(11), e1538–e1553. https://doi.org/10.1212/WNL.0000000000010256 **authors contributed equally*

Heeman, F., Hendriks, J., Lopes Alves, I., Ossenkoppele, R., Tolboom, N., van Berckel, B. N. M., Lammertsma, A. A. & Yaqub, M. O*n behalf of the AMYPAD consortium* (2020). [11C]PiB amyloid quantification: Effect of reference region selection. EJNMMI Research, 10(1), 123. https://doi.org/10.1186/s13550-020-00714-1

de Vries, B. M., Golla, S. S. V., Ebenau, J., Verfaillie, S. C. J., Timmers, T., **Heeman, F.**, Cysouw, M. C. F., van Berckel, B. N. M., van der Flier, W. M., Yaqub, M. & Boellaard, R. *For the Alzheimer's Disease Neuroimaging Initiative* (2021). Classification of negative and

positive 18F-florbetapir brain PET studies in subjective cognitive decline patients using a convolutional neural network. European Journal of Nuclear Medicine and Molecular Imaging, 48(3), 721–728. https://doi.org/10.1007/s00259-020-05006-3

Lopes Alves, I., **Heeman, F.**, Collij, L. E., Salvadó, G., Tolboom, N., Vilor-Tejedor, N., Markiewicz, P., Yaqub, M., Cash, D., Mormino, E. C., Insel, P. S., Boellaard, R., van Berckel, B. N. M., Lammertsma, A. A., Barkhof, F. & Gispert, J. D. (2021). Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging. Alzheimer's Research & Therapy, 13(1), 82. https://doi.org/10.1186/s13195-021-00819-29.

Heeman, F., Hendriks, J., Lopes Alves, I., Tolboom, N., van Berckel, B. N. M., Yaqub, M. & Lammertsma, A. A. (2021). Test-Retest Variability of Relative Tracer Delivery Rate as Measured by [11C]PiB. Molecular Imaging and Biology, 23(3), 335–339. https://doi.org/10.1007/s11307-021-01606-z

Gispert, J.D.,* Lopes Alves, I.,* Buckley, C., **Heeman, F.**, Bullich, S., Salvadó, G., Yaqub, M., Shekari, M., Altomare, D., Foley, C., Gray, K.R., Palombit, A., Manber, R., Wolz, R., Hill, D.L., Mett, A., Collij, L.E., Markiewicz, P., Jiao, J.J., Battle, M., Banton, N., Cardoso, M.J., Saint-Aubert, L., Boellaard, R., Garibotto, V., van Berckel, B.N.M., Molinuevo, J.L., Frisoni, G.B., Stephens, A., Farrar, G., Schmidt, M., Lammertsma, A.A. & Barkhof, F. *On behalf of the AMYPAD Consortium*. β-amyloid PET Imaging in the AMYPAD Project: Implementing Quantitative Imaging in Clinical Trials and Clinical Practice. *Submitted.* **authors contributed equally*

Collij, L.E., Salvadó, G., Wottschel, V., Schoenmakers, P., **Heeman, F.**, Aksman, L., Wink, A.M., van Berckel B. N.M., van de Flier, W. M., Scheltens, P., Visser, P.J., Barkhof, F., Haller, S., Gispert, J.D & Lopes Alves, I. Data-driven evidence for three distinct trajectories of amyloid-β accumulation. *Submitted.*

Heeman, F., Yaqub, M., Hendriks, J., van Berckel, B. N. M., Collij, L. E., Gray, K.R., Manber, R., Wolz, R., Garibotto, V., Wimberley, C., Ritchie, C., Barkhof, F., Gispert, J.D., Vállez García, D., Lopes Alves, I. & Lammertsma, A.A. *On behalf of the AMYPAD Consortium*. Impact of cerebral blood flow and amyloid load on SUVR bias. *Submitted.*

Jiao, J.J., **Heeman, F.**, Dixon, R., Wimberley, C., Lopes Alves, I., Gispert, J.D., Lammertsma, A.A., van Berckel, B.N.M., da Costa-Luis, C., Markiewicz, P., Cash, D., Cardoso, M.J., Ourselin, S., Yaqub, M. & Barkhof, F. NiftyPAD – Novel Python Package for Quantitative Analyses of Dynamic PET data. *Submitted.*

Legdeur, N., Moonen. J.E., Badissi, M., **Heeman, F.**, Collij, L.E., Sudre, C.H., Pelkmans, W., Gordon, M.F., Novak, G., Barkhof, F., van Berckel, B.N.M., Scheltens, P., Peters, M., Visser, P.J., & Muller M., Blood pressure in relation to volumetric MRI measures of brain aging in individuals aged 90 years and older. *Submitted.*

Acknowledgments / Dankwoord

De afgelopen jaren van zijn voorbij gevlogen. Het is een bijzonder traject geweest, waarin ik veel heb geleerd en mooie ervaringen heb opgedaan. Voor een buitenstaander lijkt een PhD-traject wellicht een eenzame route en, ik moet toegeven, dat is het af en toe ook. Maar gelukkig komt er ook veel samenwerking bij kijken. Het tot stand komen van dit proefschrift is immers niet alleen mijn werk is geweest, maar ook een team effort. Een aantal mensen wil ik daarom graag bij naam noemen en bedanken voor hun bijdrage.

Beste Adriaan, toen ik je ontmoette, was ik er nog niet 100% zeker van of ik als PhD kandidaat of als data-analist wilde werken. Achteraf gezien is het een goede keuze geweest om voor het PhD-traject te gaan. Je vriendelijke, open houding stelde mij meteen op mijn gemak. Daarbij waardeer ik je kritische blik en de manier waarop je verschillende kanten van een probleem belicht, maar de uiteindelijke keuze vrijwel altijd bij mij neerlegt. Ik vind het bewonderenswaardig van hoeveel verschillende disciplines je verstand hebt, en ik ben blij dat ik zoveel van je heb kunnen leren de afgelopen jaren. Natuurlijk zal ik naast je kennis ook de gezellige gesprekken en mooie verhalen over vroeger gaan missen.

Beste Maqsood, zonder jou was ik sowieso vastgelopen tijdens de ontelbare simulaties die ik in de eerste twee jaar van mijn PhD heb moeten uitvoeren. Ik ben heel blij met je onuitputtelijke geduld en behulpzaamheid bij het schrijven van MATLAB-programma's om mijn werk makkelijker te maken. Daarnaast waardeer ik het enorm dat je me zo vaak hebt geholpen bij het oplossen van problemen en alle goede gesprekken die we hebben gehad. Bedankt voor alles!

Dear Isadora, I am very grateful that you also joined my PhD team, although it was not from the very start. I sincerely appreciate your efforts to help me improve my writing-skills and push me just a little when you knew I could bring my work to a higher standard. This definitely helped me grow, both as a writer and as a researcher. In addition to the work-related aspects, I also really enjoyed being able to share the fun part of working in science, through visiting conferences and sharing apartments during these trips.

All opponents/members of the "reading committee", Frederik, Michel, Gitte, Rik, Valentina and Daniëlle, thank you for the time you invested in reading and assessing my thesis. In het bijzonder Frederik, bedankt zowel voor het op je nemen van de taak als voorzitter van de leescommissie, maar ook voor het mogelijk maken van mijn PhD

als onderdeel van het AMYPAD-project. Ik waardeer je interesse in hoe het me verging op het technische project waaraan ik begonnen was, je kritische vragen tijdens mijn PhD-traject, en je bereidheid om mee te denken over vervolgstappen in mijn carrière.

Thanks to all coauthors for your contributions that resulted in this thesis!

Lieve Corinne, 4,5 jaar lang hebben wij naast elkaar mogen werken als collega's. Al vanaf het eerste moment was er een klik, want jij had ook Neurosciences gestudeerd! We hebben de afgelopen jaren een hoop meegemaakt: honderden lunches, Agrafia momentjes, bruiloften, veel geklets, prachtige acteerwerk voor promotiefilmpjes, geborrel en natuurlijk ook je nuttige advies over letterlijk van alles. Dankzij jouw aanwezigheid heb ik een geweldige tijd gehad, en ik ben heel blij dat jij één van mijn Paranimfen wilt zijn.

Beste Janine, in 2019 begon je als stagiair aan een bachelor-project bij Maqsood en mij, en gelukkig ben je sinds dat project niet meer van de afdeling vertrokken. We merkten vanaf het begin dat je niet alleen kritisch, intelligent en snel was, maar óók een heel fijn persoon om mee samen te werken. Ik waardeer al het werk dat jij de afgelopen jaren als stagiair en onderzoeksassistent hebt gedaan enorm, en hoop dat je na het afronden van je Master ook een fijne, inspirerende plek vindt om te werken.

Ook alle collega's van 1F enorm bedankt voor alle gezelligheid! Matthijs, dat je mij scherp en to the point hield ("mand"), ons kantoor opvrolijkte met de "wall of fame", de vrijdagmiddagborrel nieuw leven in blies, en af en toe de zorg op je nam voor mijn kat, Tana. Thomas, voor je hulp bij alle MATLAB-issues, het kweken van enorme avocadoplanten, je droge humor, geweldige cappuccino's en al je video-edit skills voor promotiefilmpjes. Die doen zeker niet onder voor de Hollywood standaard. Yvonne, voor het opdoen en delen van de meest exotische reisideeën (Madagaskar), je hulp bij issues waar je als beginnende PhD-kandidaat tegen aanloopt (hoe beantwoord je op een nette manier de vragen van je reviewers) en je altijd welgemeende interesse. Bart, voor de gezelligheid in ons kantoor, goede discussies en natuurlijk de mooie souvenirs!

Daarnaast natuurlijk ook alle andere collega's waarmee ik een periode de F-vleugel heb mogen bewonen: Gem, Dennis, Coreline, Sanne, Marijke, Jessica, Cemile, Simone, Hayel, Sandeep, Eline, Han, Mette, Keiichi, bedankt voor de mooie tijd, de heerlijke lunches, de oneindig lange discussies, verhalen over het trechterbekervolk, momenten dat ik besefte dat ik beter had moeten nadenken voordat ik mijn mond opendeed, borrels en tochtjes

over de Amsterdamse grachten! Natuurlijk ook de 1F-aanhang, Ramsha voor je altijd vrolijke en gezellige gesprekken en het mogen meemaken van je bruiloft en Bernard voor de boottochtjes en je Polderse-steun tijdens de vele lunch discussies.

Het AMYPAD-team en alle AMYPAD-deelnemers, bedankt! Zonder jullie was er geen onderzoek mogelijk geweest. A special thanks to Yvonne, Ilona, Ilse, Ifrah and David for your contributions to the project and for offering a helping hand when I needed one. Lyduine, als co-PhD-kandidaat op het AMYPAD-project hebben we samen succesvol aan verschillende projecten gewerkt en heel wat geweldige congressen en jaarlijkse AMYPAD-bijeenkomsten meegemaakt de afgelopen jaren, bedankt! Furthermore, I would like to thank everyone from the AMYPAD team abroad and specifically Gemma, JD, Gill, Catriona, Pawel, Jieqing and Santi, for your support, collaborations and great conversations.

Alle andere RNG-collega's, Ronald, Marc, Silvia, Viktor, thank you for the good conversations and great times! In het bijzonder Alle-Meije, voor je helpende hand aan het begin van mijn PET-carrière. Alle laboranten, bedankt voor het zorgvuldig verzamelen van de scans aan de hand van het zo 'geliefde' coffee-break protocol en het secretariaat voor jullie ondersteuning de afgelopen jaren. Els en mijn mede ONWAR Neurosciences PhD-kandidaten, bedankt voor de gezellige tijd en jullie hulp bij het organiseren van de jaarlijkse ONWAR-bijeenkomst!

A special thanks also to Ines. Following our Masters' degree in Neurosciences we both decided to pursue a PhD, affiliated to the VU University Amsterdam, although you moved to Miami. Through a beautiful coincidence, one of my yearly conferences happened to be in Miami. This meant I could visit you every year and you took the time to show me all the best parts of Miami, from beaches to roof top bars, on a bike, by car or metro. I sincerely appreciate all the time and effort you have put into giving me the most wonderful experience and the great time we had together. Also, sharing a room on the annual ONWAR conference and enjoying the party nights were a great opportunity to catch up. I guess we have had our best and most creative conversations in the middle of the night. Our distant-friendship means a lot to me and I hope that even though you live in Chicago and I will move the Sweden, it will stay the same

Of course, I would not have enjoyed this PhD-trajectory as much as I did, if I had not been part of many great conferences in different parts of the world. Thanks to all Dutch

and foreign conference buddies, for making these times not just scientifically interesting but also fun and memorable for a lifetime.

Ik wil mijn vrienden bedanken voor alle gezelligheid, goede gesprekken en ontspanning (borrels, koffietjes, weekendjes weg, wandelingen en, tennis!), die als welkome afleiding kwamen in het PhD-bestaan en voor alle steun op momenten dat het even wat lastiger was.

In het bijzonder Melissa, we kennen elkaar al lang en hebben de afgelopen jaren heel wat meegemaakt en goede gesprekken gehad over onze interesses, kwaliteiten en carrière keuzes. Ik vind het erg leuk dat jij, zonder dat je van tevoren precies wist wat het inhield, één van mijn Paranimfen wil zijn.

Natuurlijk ook mijn ouders en zussen, Linda & Cindy bedankt voor jullie steun en interesse de afgelopen jaren. Het terugkeren naar mijn *roots* op de boerderij in Flevoland was een welkome afleiding van het drukke Amsterdamse leven. Lin, bedankt voor de tijd die je hebt gestoken in het checken van mijn dankwoord!

Tot slot natuurlijk de allerliefste Remy. Van tevoren was je er een beetje bang voor dat ik 4 jaar opgeslokt zou worden door mijn PhD, maar dat viel gelukkig (op sommige momenten na) reuze mee. Dat was maar goed ook, want ik had voor geen goud onze avonturen van de afgelopen jaren willen missen, die soms zelfs dankzij mijn PhD (Japan!), zijn ontstaan. Natuurlijk waren er ook lastigere momenten de afgelopen jaren, vooral de laatste loodjes van mijn proefschrift vielen zwaar. Maar als rots in de branding bewaarde jij de rust en zorgde je ervoor dat ik me niet helemaal gek liet maken. Op een toekomst met nog heel veel avonturen samen.